Automated video-based measurement of eye closure using a remote camera for detecting drowsiness and behavioural microsleeps

A thesis submitted in partial fulfilment of the requirements for
the degree of Masters of Engineering

in

Electrical and Computer Engineering
in the
University of Canterbury
Christchurch, New Zealand

by

Amol M. Malla

September 2008

Table of Contents

Acknowledgements

I have utmost gratitude to my supervisors Assoc. Prof. Richard Jones, Prof. Philip Bones, Dr. Paul Davidson, and Dr. Richard Green for supporting me throughout my thesis with their knowledge and encouragements. It was a privilege to work under their guidance. I sincerely thank my senior supervisors Assoc. Prof. Richard Jones and Prof. Philip Bones for continually challenging me to do better, whilst providing me with their invaluable advice, support, encouragement, and understanding through tough times. This report would not have been completed without their tireless effort to review and provide timely feedback despite their busy schedules. I am also very grateful to my supervisors Dr. Paul Davidson and Dr. Richard Green for their expert technical advice and guidance throughout the development phase of this project.

I express my sincere appreciation to *Christchurch Neurotechnology Research Programme* for financially supporting me with the Scholarship. My gratitude goes out to each of the research subjects, Dr. Paul Davidson, Dr. Malik Peiris, Inn Sze Low, Dr. Dong-hwan (Brian) Ko, Dr. LiPyn Leow, Dr Marcus Heitger, Ulrike Witte, and Saskia van Stockum, who gave their time and effort to record valuable reference video data. I would like to thank Dr. David Goode at Medical Physics & Bioengineering, Christchurch Hospital for his assistance with illumination measurement equipments. My sincere thanks also go to Dr. Chew Theam Yong and Govinda Poudel for their advice and help in implementing corrections in this report.

I was fortunate to have worked in company of inspirational and friendly bunch of fellow students and staffs at the Van deer Veer Institute. I would like to thank Dr. LiPyn Leow for her constant encouragement, support, and inspiration in both my academic and personal life. She will always be a true friend. I feel fortunate to have the most kind and inspirational “desk buddy”, Saskia van Stockum right through this journey. I also thank my fellow colleagues and friends Dr. Malik Peiris and Govinda Poudel for sharing ideas on different aspects of the project.

Words cannot justify my gratitude to my family. I am most thankful for the unconditional love of my parents, Achyut and Yashoda Malla, who have also been key source of inspiration, encouragement, and strength in my life. I shall forever be grateful to my late father, who passed away during this project, for exemplifying the characteristic of a good honest man and teaching me to work hard and be humble. I dedicate this work to my father.

Abstract

A device capable of continuously monitoring an individual's levels of alertness in real-time is highly desirable for preventing drowsiness and lapse related accidents. This thesis presents the development of a non-intrusive and light-insensitive video-based system that uses computer-vision methods to localize face, eyes, and eyelids positions to measure level of eye closure within an image, which, in turn, can be used to identify visible facial signs associated with drowsiness and behavioural microsleeps.

The system was developed to be non-intrusive and light-insensitive to make it practical and end-user compliant. To non-intrusively monitor the subject without constraining their movement, the video was collected by placing a camera, a near-infrared (NIR) illumination source, and an NIR-pass optical filter at an eye-to-camera distance of 60 cm from the subject. The NIR-illumination source and filter make the system insensitive to lighting conditions, allowing it to operate in both ambient light and complete darkness without visually distracting the subject.

To determine the image characteristics and to quantitatively evaluate the developed methods, reference videos of nine subjects were recorded under four different lighting conditions with the subjects exhibiting several levels of eye closure, head orientations, and eye gaze. For each subject, a set of 66 frontal face reference images was selected and manually annotated with multiple face and eye features.

The eye-closure measurement system was developed using a top-down passive feature-detection approach, in which the face region of interest (fROI), eye regions of interests (eROIs), eyes, and eyelid positions were sequentially localized. The fROI was localized using an existing Haar-object detection algorithm. In addition, a Kalman filter was used to stabilize and track the fROI in the video. The left and the right eROIs were localized by scaling the fROI with corresponding proportional anthropometric constants. The position of an eye within each eROI was detected by applying a template-matching method in which a pre-formed eye-template image was cross-correlated with the sub-images derived from the eROI. Once the eye position was determined, the positions of the upper and lower eyelids were detected using a vertical integral-projection of the eROI. The detected positions of the eyelids were then used to measure eye closure.

The detection of fROI and eROI was very reliable for frontal-face images, which was considered sufficient for an alertness monitoring system as subjects are most likely facing straight ahead when they are drowsy or about to have microsleep. Estimation of the y-coordinates of the eye, upper eyelid, and lower eyelid positions showed average median errors of 1.7, 1.4, and 2.1 pixels and average 90^{th} percentile (worst-case) errors of 3.2, 2.7, and 6.9 pixels, respectively (1 pixel $\simeq$ 1.3 mm in reference images). The average height of a fully open eye in the reference database was 14.2 pixels. The average median and 90^{th} percentile errors of the eye and eyelid detection methods were reasonably low except for the 90^{th} percentile error of the lower eyelid detection method. Poor estimation of the lower eyelid was the primary limitation for accurate eye-closure measurement.

The median error of fractional eye-closure (EC) estimation (i.e., the ratio of closed portions of an eye to average height when the eye is fully open) was 0.15, which was sufficient to distinguish between the eyes being fully open, half closed, or fully closed. However, compounding errors in the facial-feature detection methods resulted in a 90^{th} percentile EC estimation error of 0.42, which was too high to reliably determine extent of eye-closure. The eye-closure measurement system was relatively robust to variation in facial-features except for spectacles, for which reflections can saturate much of the eye-image. Therefore, in its current state, the eye-closure measurement system requires further development before it could be used with confidence for monitoring drowsiness and detecting microsleeps.

Preface

This thesis conforms to the referencing style recommended by the American Psychological Association Publication Manual (5th Ed.).

The technical aspects of this research were carried out between March 2005 and June 2007 in the Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand. The thesis was written (part-time) between June 2007 and September 2008. The project was supervised by Assoc. Prof. Richard Jones[1, 2], Prof. Philip Bones[1, 2], Dr. Paul Davidson[1], and Dr. Richard Green[1, 3]. The project was funded by a Scholarship from the Christchurch Neurotechnology Research Program.

Aspects of work from this project were presented as follows:

- "*Video-based metrics for detection of drowsiness and lapses: an overview"*; presented at Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand; 9th August 2005.

- "*Video-based metrics for detection of drowsiness and lapses: an overview"*; presented at Medical Physics and Bioengineering, Christchurch, New Zealand; 22nd November 2005.

- *"Eye detection and eye-metrics measurement using template matching and image intensity gradient analysis"*; presented at Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand; 7th February 2006.

- *"Eye localisation and eye metrics measurement using template matching and image intensity analysis"*; presented at Medical Physics and Bioengineering, Christchurch, New Zealand; 23rd May 2006.

- *"Development on video based eye closure detector"*; presented at Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand; 17th October 2006.

- *"Video-based automated eye closure detection system: results"*; presented at Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand; 8th May 2007.

1. Christchurch Neurotechnology Research Programme
2. Department of Electrical and Computer Engineering, University of Canterbury
3. Department of Computer Science and Software Engineering, University of Canterbury

Abbreviations

BM	Behavioural microsleep
CC	Correlation coefficient matrix
CNRP	Christchurch Neurotechnology Research Programme
COE_a	Manually annotated centre of eye
COE	Centre of eye (stationary centre of visible parts of eyeball; cf. sclera and pupil)
D_{close}	Distance between the centre of fROI and UEL_y when eyes fully closed
D_{open}	Distance between the centre of fROI and UEL_y when eyes fully open
EC	Fractional eye closure
elROI	Eye region of interest optimized for eyelid detection
eROI	Eye region of interest optimized for COE detection
$eROI_f$	Anthropometric proportional constants of eROI relative to fROI
fROI	Face region of interest
h	Height of the open portion of an eye.
$\hat{H}_{ant}$	Reference height of a fully open eye in the annotated reference database
$\hat{H}$	Reference height of a fully open eye
IRED	Infrared emitting diode
Lapse	Lapse of responsiveness
LEL_y	y-coordinate of apex of lower eyelid
NIR	Near infrared
OILF	Optical infrared lowpass filter
PERCLOS	Percentage of eyelids closure over time (Mallis, 1999)
SD	Standard deviation
T	An eye template image
UEL_y	y-coordinate of apex of upper eyelid
UEL_{y_open}	Reference y-coordinate of UEL_y when eyes fully open
VIP	Vertical integral projection
VIP'	Gradient of vertical integral projection

Chapter 1 Introduction

1.1 Motivation

Long and irregular hours in some occupations can lead to sleep deprivation and reduced levels of alertness during work (Huang, 2001; Marcus & Loughlin, 1996). Lapse of responsiveness at the wrong moment in high risk jobs such as commercial truck and bus drivers, pilots, air-traffic controllers, and medical house-staffs can lead to disastrous consequences, including multiple fatalities. For example, in USA, the National Transportation Safety Board estimates that 31% of commercial driver deaths during work and 58% of commercial truck accidents annually can be attributed to driver fatigue (NTSB, 1995). There have also been reports of shipping disasters and train accidents due to reduced alertness of operators from sleep deprivation (Kolstad, 1990; Trosvall & Akerstedt, 1987).

Drowsiness is considered one of the main factors in private motor vehicle accidents (Lal & Craig, 2001; Stutts et al., 1999). For example, a study by American Automobile Association Foundation for Traffic Safety, USA, considered the driver drowsiness to be the second major reason, after alcohol, for car accidents (Stutts et al., 1999). The National Highway Traffic Safety Administration (NHTSA), USA, estimates that approximately 100,000 police-reported traffic accidents annually can be attributed to driver drowsiness or fatigue. These traffic accidents account for 1,200 deaths and 76,000 injuries annually in USA alone (Rau, 1996). NHTSA estimates a financial cost of US$12.5 billion each year due to drowsiness-related accidents (J. S. Wang et al., 1996). In France fatigue has been estimated to account for 10% of serious traffic accidents (Philip et al., 2001). Similarly, fatigue has been estimated to account for 20% of motorway traffic accidents in UK (Horne & Reyner, 1995).

1.2 Project background

A device capable of continuously monitoring an individual's levels of alertness and able to detect the onset of lapses in real-time is highly desirable for preventing drowsiness and lapse related accidents. The Christchurch Neurotechnology Research Programme (CNRP) has established a research programme for "detection and prediction of drowsiness and lapses of

consciousness" to understand the underlining phenomenon of lapses and its application to countermeasure drowsiness and lapse related accidents. Investigations of electroencephalogram (EEG), electrooculogram (EOG), visuomotor tracking behaviour, and subjective video-based assessment metrics have identified important characteristics of lapses (Davidson et al., 2007; Peiris et al., 2006a). Ultimately, the research programme aims to automate the measurement of all metrics of drowsiness and lapses and combine these metrics to develop an automated real-time warning device to prevent accidents. Ideally, the system would also be able to predict the likelihood of imminent lapses.

Within CNRP, Peiris et al. (2006a) determined the rate and characteristics of lapses of responsiveness in 15 non-sleep-deprived healthy subjects performing a visuomotor continuous tracking task (CTT) during two 1 hr sessions. The CCT was designed to somewhat simulated extended monotonous driving. In the study, full-head EEG, EOG-based eye movements, tracking kinematics behaviour, and facial video were also recorded simultaneously.

To date, various signal processing and subjective analyses have been applied to the acquired data in order to reliably identify high-temporal-resolution data patterns during lapses (Davidson et al., 2007; Peiris et al., 2004; Peiris et al., 2006b; Peiris et al., 2006a; Peiris et al., 2005a). In particular, a subcategory of lapses called behavioural microsleeps (BMs) has been conservatively identified in the visuomotor tracking behaviour data and video recordings. The behavioural changes observed in the video data were also used for subjective video rating of alertness levels by an expert[1] (Peiris et al., 2006a).

Subjective assessment of the video recordings was considered the most conservative indicator of BMs (Davidson et al., 2005). Ideally, video-based subjective rating of alertness levels and identification of BMs would be performed by multiple raters. However, the subjective rating and detection of BMs in 30 hr of video data from the 15 subjects by a single expert was an extremely time consuming task, which discouraged further visual rating of video data by other expert raters. It was also noted that the subjective assessment of the video data by multiple raters would highly likely introduce substantial inter-rater variability due to the subjective, monotonous, and time-consuming nature of the rating task (Davidson et al., 2005).

As an alternative to subjective rating, the development of an automated real-time video-based drowsiness and lapse-related facial metrics detection system would save a considerable amount

[1] One of the investigators (M.T.R Peiris) subjectively rated the levels of alertness and BM from the recorded video data for all 15 subjects.

of time and provide quantitative measurement for objective assessment. In addition, the integration of objective video-based metrics with EEG and tracking behaviour kinematics based metrics would improve the reliability of the ultimate multi-modality lapse detection system. An automated video-based facial metrics detection system could also be used as a stand-alone unit for monitoring drowsiness and detecting lapses.

This Master's project was initiated with the aim to develop an automated video-based system to detect drowsiness and lapses by automatically measuring the relevant facial metrics in digital images. This thesis presents the system design and computer-vision algorithms developed to automatically detect the position of the face, the eyes and feature of the eye in a digital image so as to measure the facial metrics associated with drowsiness and lapses.

Automatic extraction of facial metrics related to alertness levels from the digital video data is not novel (Grace et al., 1998; Heitmann et al., 2001; Makito et al., 1997; Ueno et al., 1994; Zhu & Ji, 2004b). Also, there are commercial products available for measuring face and eye metrics which can be used for detecting BMs (see Chapter 2). However, the development of an in-house automated video-based system would allow the CNRP's specific requirements to be integrated into the system design. In future, an in-house system would also facilitate convenient system maintenance and modification to accommodate any changes in requirements. Computer-vision algorithms for measuring facial metrics derived in this project could also be used in future research project in the CNRP. An in-house automated video-based system would also confer ownership and flexibility to develop new lapse detection methods as understanding of the phenomenon of lapse increases.

1.3 Basic surface anatomy of the eye

The eyes are the most distinctive visual feature on the face and provide a good indicator of drowsiness and BMs. The visible structure and behavioural characteristics of an eye contains many visual cues, such as colour, shape, contrast, texture, and movement patterns that can be utilized to detect it in an image. The visible parts of an eye comprise the eyeball and superficial accessory structures (Tortora & Grabowski, 2003) as shown in Figure 1-1.

The accessory structures comprise the eyebrow, eyelids, eyelashes, and lacrimal caruncle, which, in turn, form other feature such as eyelid fold and eye corners. The colour, shape, and

thickness of the eyebrows vary widely, with males usually having thicker eyebrows than female.

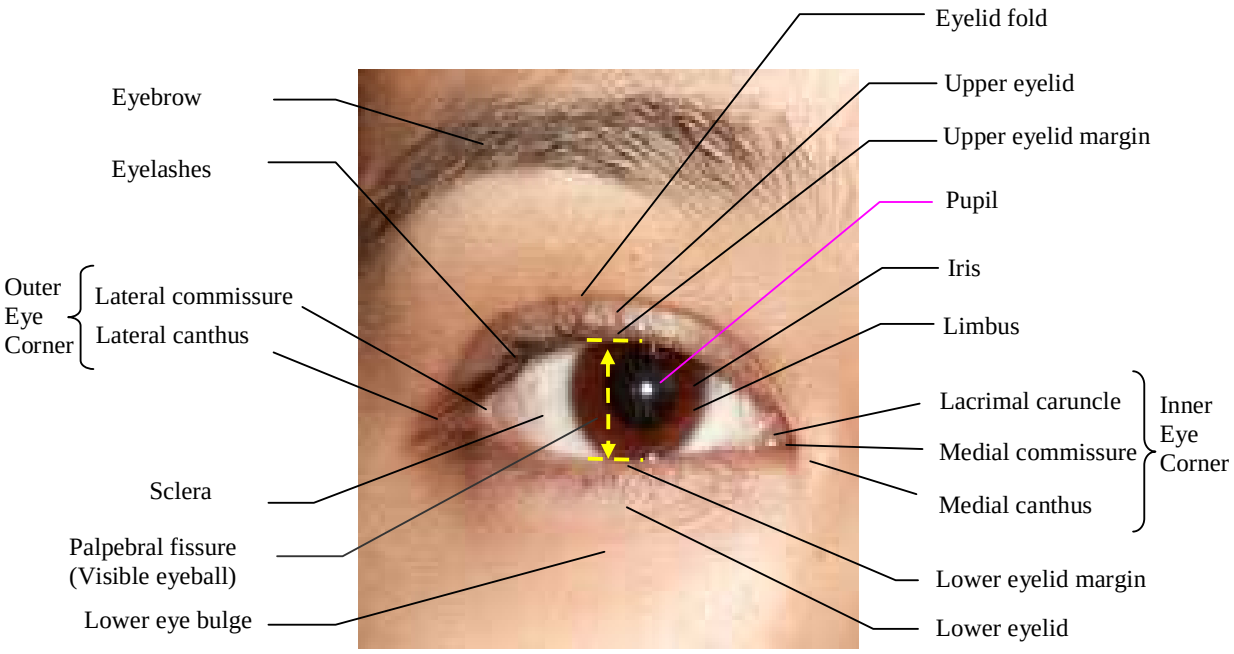


Figure 1-1. Photo of a right eye labelled with the visible surface eye anatomy.

The upper and lower eyelids are the most distinct and expressive eye feature of drowsiness. They are also anatomically known as superior and inferior palpebrae, respectively. The eyelids block out light during sleep, protect the eye from excessive light and foreign objects, and spread tears over the eyeball for lubrication and moisture. The upper eyelid is supported by specialised retractor muscles that produce its movements. Conversely, the lower eyelid has no specialised retractor muscle and its movements are minimal (Tortora & Grabowski, 2003). Movement of the upper eyelid is a combination of rotational and vertical motions (Evinger et al., 1991); it rotates as it follows the curvature of the eyeball, while changing the vertical position. In frontal view, the arch of the open upper eyelid becomes flatter and eventually arches in the opposite direction during eye closure. As the eye opens, the upper eyelid moves towards the eyelid crease or eyelid fold. If the eyelid crease is visible when the eyes are fully open2, the eyelid is known as the double eyelid and, if not visible as in some Asian

[2] In this report, the fully open eye refers to when the eye is relaxed and naturally open without squinting or forced wide open.

populations, the eyelid is known as the single eyelid as seen on Figure 1-2 (Miyake et al., 1994).

(a) (b)

Figure 1-2. Types of eyelid folds, left: single eyelid, right: double eyelid.

Evinger et al. (1991) have measured amplitude-maximum velocity eyelid movement using electromyography (EMG) to characterize upper-eyelid movements into three basic categories: voluntary eyelid movement, blinks, and saccades. Voluntary eyelid movements include intentional eye movements such as forced eye closer and voluntary blinks. Blinks are further categorized into two types: spontaneous blinks to lubricate the eyeballs and reflexive blinks to protect the eyes from foreign object. In alert and healthy people, the spontaneous blink is defined as a rapid (approximately 250 ms) closing and opening of the upper eyelid. A healthy adult blinks spontaneously on average 20 times/min although this rate changes with attention, stress, mood, eye irritation, fatigue, and amount of sleep (Barbato et al., 1995; Karson, 1992; Karson et al., 1981; Karson et al., 1986; Van Orden et al., 1998). The upper eyelid also moves with eye saccades. It rises slightly during an upward saccade and falls with a downward saccade.

The upper and lower eyelids meet to form the medial and lateral commissure. In this thesis, the area containing the lacrimal caruncle, medial commissure, and the medial canthus are collectively referred to as the “inner eye corner”. Similarly, the area containing the lateral commissure and canthus are referred to as “outer eye corner”.

The inner edges of the eyelids where they meet the eyeball are called eyelid margins. The eyelashes lie outside the eyelid margins. The space between the upper and lower eyelids that exposes the eyeball is called the palpebral fissure. In this thesis, the palpebral fissure is referred to as the “visible eye”. About one-sixth of the surface area of the eyeball is visible through the palpebral fissure when the eyelids are fully open (Tortora & Grabowski, 2003).

The visible part of the eyeball is divided into the dark coloured region of the iris and the surrounding white region of the sclera (Tortora & Grabowski, 2003). The cornea is a transparent dome-shaped coating that covers and protects the coloured iris and helps focus the light into the eye. The iris is the circular coloured portion of the eyeball consisting of radial and circular muscles. The iris regulates the amount of light entering the eye and dilates its central circular aperture called the pupil. The iris constricts to decrease the diameter of the pupil in bright visible light to restrict the amount of light entering the eye and dilates to increase the diameter of the pupil under lower visible light levels to let more light to reach the eye's photoreceptors. The colour of the iris varies considerably between individuals. The outer boundary of the iris, where it meets the sclera, forms a dark ring (regardless of the colour of the iris) called the limbus. The sclera covers and protects the rest of the eyeball, providing its rigidity and shape.

A basic sagittal section of the eyeball is shown in Figure 1-3. The inner lining of the eyeball is called the retina. It contains photoreceptors that sense visible light and transfer their signals via the optic nerve to the occipital visual cortical region of the brain. The retinal photoreceptors are sensitive to light within the visible electromagnetic spectrum of approximately 400 nm to 700 nm long wavelengths (Kandel et al., 1991).

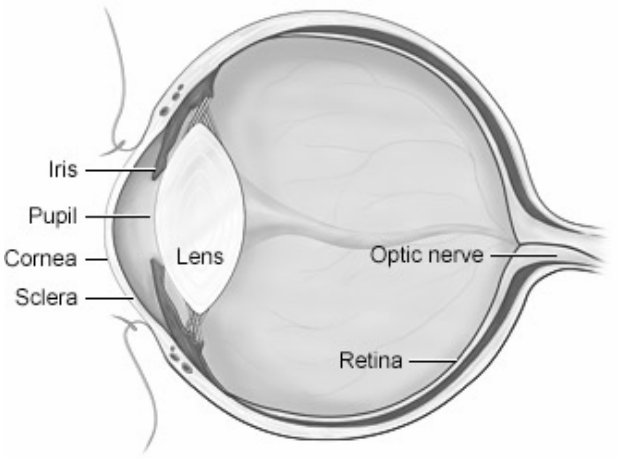
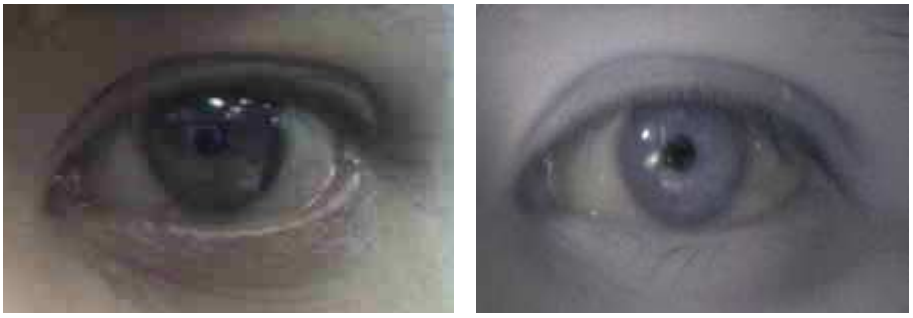


Figure 1-3. Basic internal eye anatomy (http://www.lehp.org/diseases.htm).

1.4 Eye under infrared illumination

The electromagnetic spectrum within the 730 nm to 2500 nm wavelength is known as the near-infrared (NIR) spectrum (Raghavachari, 2000). NIR is invisible to the human eye but can freely enter the pupil. Under NIR lighting, the iris and limbic boundary has lower intensity in an image formed by a camera sensitive to NIR. Figure 1-4 shows an iris of a same person under normal ambient fluorescent lighting conditions and under NIR (880 nm) spectrum.



(a) (b)

Figure 1-4. Change in appearance of the same iris under two different lighting conditions: (a) the iris appears darker under ambient fluorescent light and (b) lighter in NIR spectrum.

If the NIR source is placed at the optical axis of the camera, the pupil appears bright in the image, as shown in Figure 1-5. The pupil appears bright because the majority of NIR that reaches the retina is reflected back through the pupil.

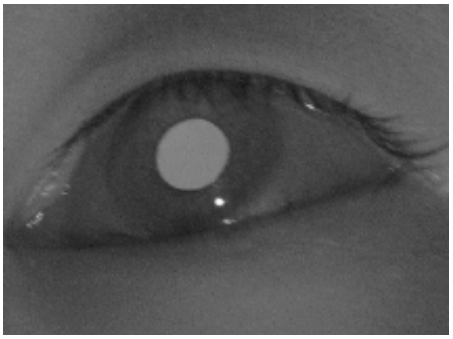


Figure 1-5. Bright pupil effect due to NIR retinal reflection (Morimoto et al., 2000).

Extended exposure of high NIR radiation to the retina can damage the retina. High intensity of NIR is more likely to damage the retina than visible light because the protective reflex mechanism of the eye does not respond to NIR spectrum. The International Commission on Non-Ionizing Radiation Protection (ICNIRP) has set a limit on maximum infrared (780-3000 nm) permissible exposure to the eye for periods longer then 16 min to 10 mW/cm^2 irradiance (Matthes, 2000). The irradiance (mW/cm^2) at the eye is a function of the power emitted by the NIR source, the area over which that energy is spread, and the uniformity of the illumination pattern (Babcock & Pelz, 2004; Barna & Schlanger, 2004).

1.5 Types of head movement

The appearance of facial features in image depends on head orientation. Hence, it is important to define various head movements to understand the varying characteristics of the eye image. Head movements can be broken down into three axes of rotations (Cohn et al., 2003) and translational movements, as shown in Figure 1-6. The three axes of rotations are pitch, roll, and pan. The head pan is also known as yaw in some literature. Aspects of facial features disappear from the visual axis of the camera as the head rotates. Whereas, the face is fully visible during translational movement until the head moves out of the field of view of the camera.

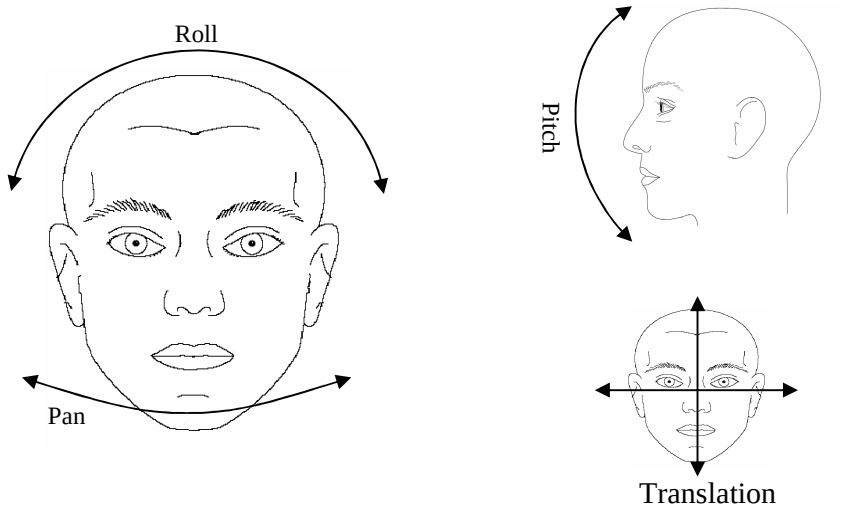


Figure 1-6. Head rotations and translational movements [face figures adapted from (Farkas, 1994)].

1.6 Lapses and microsleeps

Lapses of responsiveness (‘lapses’) are brief episodes in which an individual unintentionally stops responding to the task they are performing (Peiris et al., 2006a). A lapse is believed to be induced by temporary deactivation of attention and/or arousal neural networks in the cortical region of brain that are responsible for task performance (Davidson et al., 2007; Foucher et al., 2004; Parasuraman & Davies, 1984). Many factors can induce a lapse, including: boredom, mental-fatigue, monotonous environment, circadian rhythm, and sleep deprivation (Lal & Craig, 2001; Ogilvie, 2001; Oken et al., 2006). Lapses have also been reported in normal non-sleep-deprived subjects performing sustained attention tracking task during normal working hours (Peiris et al., 2006a).

Microsleeps are arousal-related lapses that can be identified either by EEG or behavioural signs of sleep. These two types of microsleeps are respectively called the EEG microsleep and the behavioural microsleep (Peiris et al., 2006b). EEG microsleeps are usually identified via bursts of theta activity in the brain lasting for anywhere from 1 s to 30 s (Harrison & Horne, 1996; Parasuraman & Davies, 1984). EEG microsleeps are associated with reduced level of cortical arousal and can occur without external behavioural signs of reduced arousal (Davidson et al., 2007).

Behavioural microsleeps (BMs) are defined as brief episodes of behavioural signs of sleep and/or cessation of response to the moving target while performing a visuomotor tracking task (Peiris et al., 2006a). A BM in video data is conservatively identified by subjectively observing the brief episodes of behavioural signs of sleep such as prolonged eyelid closure, sometimes accompanied by rolling upward or sideways eye movements, and head nodding that is often terminated by waking head jerks (Peiris et al., 2006a). Behavioural microsleeps related to tracking behaviour are identified by detecting the periods of “flat spots” in which the subject stops tracking the target during the visuomotor tracking task. The simultaneous occurrence of a video BM and a flat spot is defined as a definite BM. Although definite BMs were commonly observed, the video BMs and flat spots were also observed on their own during the analysis of video and tracking behaviour recordings (Davidson et al., 2007; Peiris et al., 2006a). Figure 1-7 illustrates the occurrence of flat spots, video BMs, and definite BMs from tracking behaviour and subjective video-based alertness rating data.

This thesis is particularly interested in automated detection of video-based behavioural microsleeps. Henceforth in this report, the terms “lapse”, “microsleep”, and “behavioural

microsleep" (but not EEG microsleep) are used interchangeably used to represent video-based behavioural microsleep, unless otherwise specified.

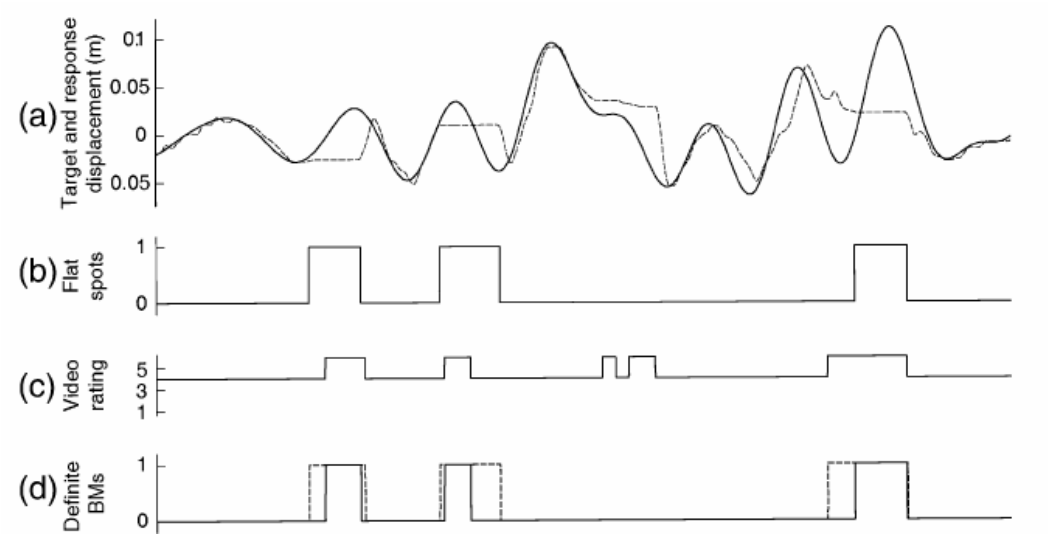


Figure 1-7. Illustration of (a) tracking behaviour while performing 1D visuomotor continuous tracking task, (b) detection of flat spots related to BM in the tracking behaviour, (c) duration of subjectively detected BM in video recording, (d) duration of definite BM derived from (b) and (c). [Adapted from Peiris et al. (2006a)]

1.7 Drowsiness

Drowsiness is simply defined as "one's tendency to fall asleep" (Shen et al., 2006). Drowsiness is a transitional physiological state between wakefulness and sleep (Oken et al., 2006). An individual can be drowsy for a prolonged period of time but not necessarily fall asleep. However, if uninterrupted, drowsiness is highly likely to progress to sleep onset process and eventually sleep (Ogilvie, 2001). Drowsiness is heavily influenced by the circadian cycle and can also be induced by factors such as sleep-deprivation (Barbato et al., 1995; Dinges et al., 1997; Marcus & Loughlin, 1996), sedative medication (Oken et al., 2006; Shen et al., 2006), monotonous environment (Horne & Reyner, 1995; Trosvall & Akerstedt, 1987), and mental fatigue (Grandjean, 1979; Lal & Craig, 2001). Sleep-related variables such as decreased cognitive and psychomotor performance, mood, motivation, autonomic and physiological changes, and microsleeps are often used as indicators of drowsiness (Dinges et al., 1997; Lal & Craig, 2002; Shen et al., 2006).

Drowsiness is associated with visible behavioural changes in the face and body posture. For example, a reduction or cessation of spontaneous blinks is considered one of the earliest reliable sign of deep drowsiness (Lal & Craig, 2001; Santamaria & Chiappa, 1987). In addition, behavioural changes such as reduced facial tone, droopy partial eyelid closure, and BMs are used for subjective determination of level of alertness, including level of drowsiness (Peiris et al., 2006a; Wierwille & Ellsworth, 1994a).

1.8 Facial metrics of drowsiness and microsleeps

The measurement of distinct patterns in physiological signals and the deterioration in performance of task provide a strong objective indication of drowsiness and occurrence of lapses that are otherwise externally not visible. However, distinct visible behavioural signs of sleepiness such as reduced facial tone, BMs, and head nods are natural and the strongest visible indication of drowsiness and microsleeps (Davidson et al., 2007).

In particular, facial signs of sleep have been repeatedly used for subjective assessment of drowsiness and lapses (Cohn et al., 2003; Galley & Schleicher, 2002; Lal & Craig, 2001; Morris & Miller, 1996; Oken et al., 2006; Peiris et al., 2005a; Santamaria & Chiappa, 1987; Van Orden et al., 1998, 2000; Wierwille & Ellsworth, 1994a). Frequent and fast saccadic eye movements, fast blinks, and tense head posture of an individual are indicative of the awake and alert state (Lal & Craig, 2002; McGregor & Stern, 1996; Peiris et al., 2006a). As a person becomes drowsy, distinct behavioural changes in the face can be seen. Some of these signs include rubbing of eyes to relieve eye fatigue, yawning, subdued appearance, reduced facial tone, decrease in pupil diameter (Morad et al., 2000), slow eye movements, an increase in blink duration, prolonged partial or full eyelid closure, and jerky head nods (Ji & Yang, 2001; Lal & Craig, 2001; Makito et al., 1997; Oken et al., 2006; Peiris et al., 2006a; Wierwille & Ellsworth, 1994a). The facial behavioural signs of sleepiness that have been more prominently used for detection of drowsiness and microsleeps are blink rate, amplitude, and duration, partial or full eyelid closure, slow eye movement, fixated gaze, and head nods.

1.8.1 Blinks and eye closure

The reduction or cessation of spontaneous blinks is considered as one of the earliest reliable sign of drowsiness (Lal & Craig, 2001; Santamaria & Chiappa, 1987). A decrease in blink

amplitude, increases in blink duration, and a decrease in blink rate have been shown to correlate with deterioration in performance on tracking tasks due to drowsiness (Morris & Miller, 1996; Van Orden et al., 2000). Changes in blink behaviour are best observed in video and EOG data. Forced eyelid closure, where a person intentionally closes their eyes to relieve eye fatigue, is another behavioural sign often associated as the early sign of drowsiness (Peiris et al., 2006a).

Degree of eye closure is one of the most important metrics for detecting drowsiness and behavioural microsleeps. Percent eye closure (PERCLOS) is a metric for estimating the level of drowsiness based on percentage of time the eyes are at least 80% closed within 1 min (Mallis, 1999; Wierwille & Ellsworth, 1994a). In a driving simulator, alert drivers are shown to have a much lower PERCLOS than the drowsy drivers (Wierwille & Ellsworth, 1994a). PERCLOS is considered the most reliable and valid visual measure of a driver's alertness by the Federal Highway Administration, USA (Ji & Bebis, 1999; Wierwille & Ellsworth, 1994b) and it has become a well adapted metric in many drowsiness monitoring systems (Grace et al., 1998; Ji & Yang, 2001; Mallis, 1999; Zhu & Ji, 2004b).

Partial droopy eyelid closure is also a good indicator of deep drowsiness (Lal & Craig, 2002; Morris & Miller, 1996; Peiris et al., 2006a). During the onset of sleep there is a general tendency for muscles, including the upper eyelid's levator palpebrae, to relax. A relaxed upper eyelid muscle results in slow eyelid closure usually perceived as droopy eyes. Finally, the detection of complete eye closure for a prolonged period of time can be used for identifying events of microsleep.

1.8.2 Eye movement

During drowsiness, the velocity of saccadic eye movements decreases (McGregor & Stern, 1996) and also gaze becomes more fixated with reduced visual scanning of the environment (Morris & Miller, 1996; Santamaria & Chiappa, 1987; Van Orden et al., 2000). Under deep drowsiness, slow rolling eye movements can also be observed (Ogilvie, 2001).

1.8.3 Head nods and orientations

Head droop and frequent nodding behaviour are one of the strongest indicators of deep drowsiness and microsleep (Lal & Craig, 2002; Peiris et al., 2006a; Santamaria & Chiappa,

1987). While studying driver fatigue in a driving simulation task, Lal & Craig (2002) found that 23% of their subjects showed an average of 1 nod every 10 s during the deep drowsy state. In a study of lapses on a tracking task by Peiris et al. (2006a) observed an average 35 video-based microsleeps per hour. These were characterized by momentary full eye closure sometimes followed by head nods usually terminated by waking head jerks. Detection of sideways, upward, or downward head orientation for a prolonged period of time can also be used to identify a distracted or inattentive driver if the upright frontal head orientation is assumed to be the nominal head orientation while driving a vehicle (Ji & Yang, 2002).

1.9 Facial features for the facial metrics

Prominent facial signs of drowsiness and microsleeps can be objectively detected by quantitative measurement of three facial metrics from the frontal face digital video of an individual: eyelid closure, slow eye movements, and head movement. The specific face and eye features that must be detected in order to derive the facial metrics associated with drowsiness and microsleep are discussed in next few sections.

1.9.1 Eyelids

Detecting the position of the eyelids in the frames of a video allows measurement of eye closure. Apex in an eyelid is vertically the furthest point in the eyelid margin (see section 1.3). Hence, estimating the y-coordinate of apex of upper eyelid (UEL_y) and lower eyelid (LEL_y) as shown in Figure 1-8 will allow the measurement of eye closure. Measurement of eye closure can be used for identifying various behavioural facial signs of drowsiness and microsleeps such as, the forced eye closure, prolonged partial or full eyelid closure, reduced blink rate and amplitude, and increased blink duration.

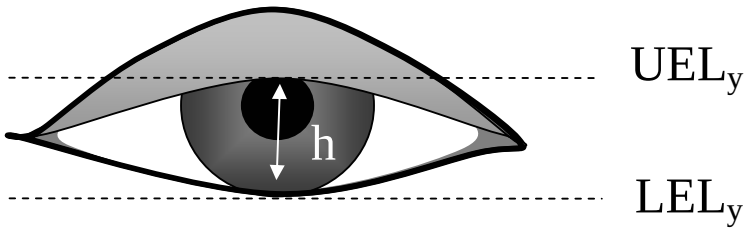


Figure 1-8. Apex of upper and lower eyelids must be detected to measure eye closure measurement.

1.9.2 Pupil and eye corners

The estimation of centre of pupil's position (p) relative to a fixed reference point in a face can be used to detect slow eye movement (i.e., no saccades) and fixated gaze, which are two good indicators of drowsiness. Also, a gaze direction away from straight ahead gaze for a prolonged period of time is indicative of distracted driver. A fixed reference point (C) can be estimated by detecting and averaging the coordinates of inner (C_i) and outer (C_o) eye corners, which are locally static features in the eye region. Hence to measure eye movement, the developed system must be able to automatically locate the position of the centre of pupil and the inner and outer eye corners, as shown in Figure 1-9.

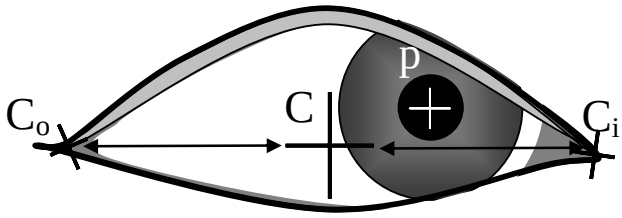


Figure 1-9. The centre of pupil (p), inner eye corner C_i and outer eye corner C_o must be detected to measure eye movement and estimate gaze direction.

1.9.3 Head position and orientation

Detecting the position of the head in consecutive frames can be used to estimate direction and velocity of the head movement, which, in turn, can be used to identify head nods and head orientation. Head nods are one of the definite signs of behavioural microsleep. They are usually characterized by a slow downward pitch movement of the head (droopy head movement) followed by fast jerky pitch movement that usually recovers to an upright head position. In its simplest form, the determination of vertical head movement velocity from position of the head in consecutive frames of the video can be used to identify head nods. Detection of head position in 3D can be used to detect head orientation. The detection of head orientation other than the frontal upright can be used to identify and warn distracted subjects in activities such as driving.

Chapter 2 Review of video-based facial feature detection methods

In recent years, video-based facial feature detection and tracking systems have been employed in a wide range of applications. These include drowsiness monitors, human-computer interfaces, biometric security systems, robotics, psychological studies, and product marketing (Duchowski, 2003). Improvements in the robustness and accuracy of computer vision methods and the availability of low-cost off-the-shelf hardware like webcams with relatively good performance has accelerated the research and development of the video-based facial feature detection and tracking systems. Recently, commercial companies producing video-based systems for facial feature detection and tracking for a wide range of applications have also been established. This chapter reviews the literature on different types of video-based face and eye feature detection systems, the computer vision methods used in these systems, and some of the commercial products currently available.

2.1 Head-rest systems

The setup of an eye detection system depends on the accuracy requirements and the user compliance of its application. For example, for research applications in a controlled laboratory environment, the system often requires high precision measurements but can allow some degree of physical constraints to the user. Eye feature tracking systems used in research applications often use head-rest units (Duchowski, 2003), which constrain movement to minimize errors due to head motion during eye tracking. The “iView X Hi-Speed” eye movement tracking device developed by SensoMotoric Instruments, Boston, U.S.A, as shown in Figure 2-1, is an example of one such commercial eye movement tracking system which is currently used in the Van der Veer Institute. The advanced version of iView system can sample at 1250 Hz with eye movement tracking accuracy of less than 0.01˚ and gaze estimation accuracy of 0.2˚ (SMI, 2005). Head-rest systems like these provide high precision measurements of the eye parameters but are not practical outside the laboratory environment because they constrain head movement.

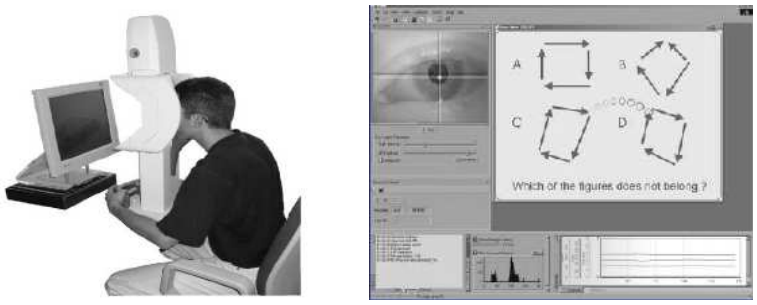


Figure 2-1. iView X Hi-Speed head-rest based gaze detection system developed by SensoMotoric Instruments (SMI, 2005).

2.2 Head-mounted systems

To make the eye tracking system more portable while minimizing noise due to head motion, head-mounted devices with a built-in camera have also been developed. A head-mounted device developed by Babcock & Pelz (2004) for gaze estimation is shown in Figure 2-2.

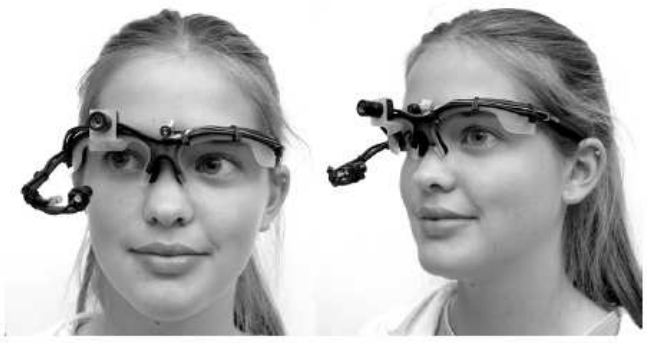


Figure 2-2. Headgear for eye tracking and gaze detection developed by Babcock & Pelz (Babcock & Pelz, 2004).

One of the advantages of the head-rest and head-mounted systems is that the camera is usually mounted close to the eye, which provides a high resolution image for precision measurements. Also, the camera mounted close to the eye will contain only the eye in its entire field of view, making it unnecessary for the use of eye localization methods. The disadvantage of head-gear systems is that they are intrusive and often impractical for use in applications such as driver drowsiness monitors, where the eye metrics have to be monitored for an extended period of time. As the primary aim of the video-based system to be developed in this project was to

detect drowsiness and microsleeps under unconstrained environment, the head-gear approach was considered impractical and not pursued.

2.3 Remote camera-based systems

A remote camera-based system is the most practical and natural solution for applications where the video-based facial metrics need to be acquired without intrusion to the subject and their operational environment. Some of the applications where remote camera-based systems seem ideal are drowsiness estimation (Grace et al., 1998; Zhu & Ji, 2004b), biometric security (Daugman, 2004; El-Bakry, 2001; Feng & Yuen, 1998) and human-computer interfacing (Bradski, ; Duchowski, 2003; Ebisawa & Nurikabe, 2006; Noureddin et al., 2005). In a remote-camera based system, subjects are often at a fixed distance from the camera with the entire face in the camera's field of view. Unlike the head-gear systems, remote camera-based system has to first localize the face and/or eye region in an image. Localization of the face region of interest reduces the search area for the eyes within the image. Once the eyes are localized, eye features such as pupil, iris, eyelids, and eye corners can be detected. The eye feature can then be used for estimating facial metrics such as eye movements, gaze direction, percentage of eye closure, and blinks.

In a given image the eyes can be localized either by an active or a passive eye detection method (Ji et al., 2005; P. Wang & Ji, 2005). Active eye detection methods utilize some form of active markers on the eye and usually do not require the localization of the face region of interest. On the other hand, most passive eye feature detection methods rely on the localization of the face and eye regions of interest to reduce the search area before detecting the eye features.

The reflective property of different parts of the eye can be used to actively mark the eye with an external light source. The retinal reflection of a near infrared (NIR) source and the corneal reflection of a light source are the two most widely used eye markers for active localization of the eyes.

2.3.1 Active eye detection methods

2.3.1.1 Retinal reflection method

If an NIR source is placed close to the optical axis of an NIR sensitive camera that is directed towards the face of a person, the image of the pupil appears relatively very bright with a darker

iris around it due to the retinal reflection of the NIR source. The NIR retinal reflection of the eye seen in an image is called the bright pupil effect (Ebisawa & Satoh, 1993) and is also known as the red eye effect. In contrast, if the NIR source is placed further away from the optical axis of the camera, the pupil image appears dark with relatively lighter iris around it. The diagram in Figure 2-3 demonstrates the setup for producing the bright and dark pupil effects.

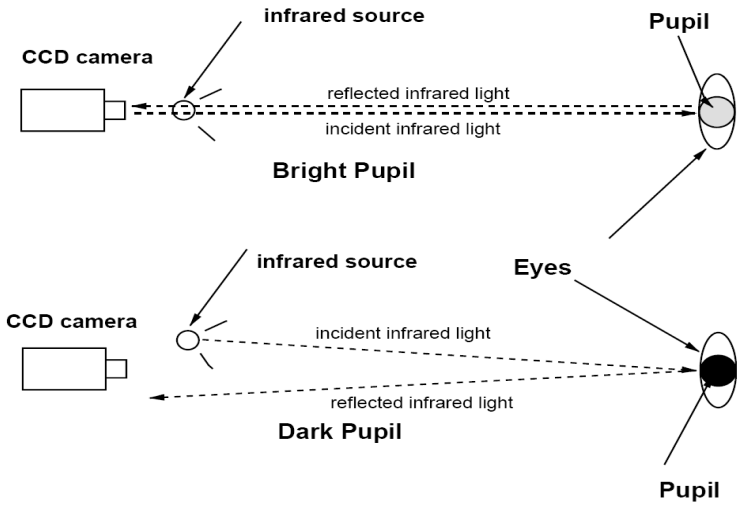


Figure 2-3. Positioning of the NIR source and camera to produce bright and dark pupil effects (Ji & Yang, 2002).

If an NIR source is placed close to the optical axis and another further away from the optical axis of a camera, the bright and dark pupil effects can be produced in the consecutive frames of the video by alternatively switching the corresponding inner and outer NIR sources synchronous to the video capture rate. Ji et al. (2004) produced the bright and the dark pupils in two consecutive frames of the video by synchronizing the frame rate to the corresponding inner and outer rings of NIR LEDs placed concentric to the optical axis of the camera. The bright and dark pupils in consecutive interlaced frames are shown in Figure 2-4 (a) and (b) respectively. The difference between consecutive frames with bright and dark pupils is then computed to form a difference image with two bright blobs at the position of the pupils as shown in Figure 2-4 (c). The subtraction of the consecutive frames cancels out the static background image and improves the signal-to-noise ratio of the bright pupil blobs. The positions of the bright pupils, and hence, the eyes, are then localized in an image by applying

intensity threshold method and *a priori* distance constraint between the centres of the pupil blobs (Ji et al., 2004; Liu et al., 2002).

a b c

Figure 2-4. (a) Bright pupil effect. (b) Dark pupil effect. (c) The image with the distinct bright blobs at the pupil locations after subtracting the frames in (a) and (b) (Ji et al., 2004).

Once the blob of the bright pupil is detected, an ellipse is fitted to the blob to extract various eye metrics (Bergasa et al., 2006; Ji et al., 2004). The change in ratio of radiuses of the ellipse due to the occlusion of the pupil by the eyelids is used to determine the percentage of eye closure. By measuring the percentage of eye closure as a function of time, the PERCLOS metric is used to detect driver drowsiness (Bergasa et al., 2006; Grace et al., 1998; Ji et al., 2004). The detection of rapid eyelid closure and the absence of bright pupils are also used for identifying blink rate and duration (Bergasa et al., 2006; Seki et al., 1998). Ji and Yang (2002) have also estimated the 3D face orientation based on change in pupil size, pupil shape, and distance between the centres of left and right pupils. The pupil positions relative to a known fixed reference marker, such as a glint or an eye corner, is also used for estimating gaze direction (Zhu and Ji 2004; Morimoto and Mimica 2005). The gaze direction away from the driver's nominal frontal gaze for a long period of time is used for identifying distracted or inattentive drivers (Bergasa et al., 2006; Ji & Bebis, 1999). The estimation of the gaze direction derived from the bright pupil effect is also used for hands-free navigation in human-computer interface systems (Ebisawa & Nurikabe, 2006; Morimoto & Mimica, 2005; Zhu & Ji, 2004a). The eye localization based on bright pupil detection has also been used for automated facial expression analysis (Kapoor & Picard, 2002).

Advantages of using retinal reflection

There are a number of advantages of using retinal reflection based active eye and eye feature detection system. The detection of the bright pupil allows direct localization of the eye position and pupil size without first localizing the face region of interest in an image. The high contrast

of the bright pupil significantly improves the robustness of the eye detection system. The NIR illumination also makes the system invariant to changing visible illumination conditions. Since the NIR is invisible to the human eye, the NIR source does not distract the subject from their task.

Disadvantages of using retinal reflection

However, there are also disadvantages of using the bright pupil effect. The performance of the bright pupil blob extraction from the image is dependent on the correct selection of the intensity threshold level (Bergasa et al., 2006; Ebisawa, 1995; Ji & Yang, 2002). However, the image intensity level of the bright pupil for a subject is dependent on many factors such as the distance between the subject and the camera, the subject's gaze direction, face orientation, pupil diameter, presence of corrective lenses, and degree of eye closure (Bergasa et al., 2006; Liu et al., 2002; Nguyen et al., 2002). For example, the image intensity of the bright pupil is dependent to the diameter of the pupil, which can significantly vary between subjects even under same illumination levels since the iris responds differently for different groups of people (Nguyen et al., 2002). The variation in image intensity of bright pupil makes it difficult to set a correct intensity threshold. The dynamic adaptive threshold setting algorithms can be applied to minimized the effects of varying pupil intensity (Bergasa et al., 2006). However, there are shortcomings with the adaptive threshold methods as well.

Bergasa et al. (2006) found that the performance for bright pupil detection dropped substantially under daylight conditions due to constriction of pupil which decreased the size of the bright pupil blobs and signal-to-noise ratio. Increasing the intensity of the NIR source will increase the intensity of the bright pupil in daylight conditions but it can also saturate the video signal making it difficult to detect bright pupils (Ebisawa, 1995). The high intensity of NIR also increases the risk of thermal damage to the retina.

Since the diameter of the pupil can vary, the precision of eye closure measured based on the occlusion of bright pupil by the eyelids will also vary. Hence, it is more appropriate to measure the eye closure based on the distance between the apex of the upper and lower eyelids an in this project. However, it can be argued that measurement of pupil occlusion is more important for monitoring alertness because vision is only impaired when the pupils are occluded.

Detection of bright pupil in difference image formed by subtracting consecutive frames also become very difficult under a large and fast head movement. In the difference image, substantial head movement produces artifacts that are of similar or higher intensity than the

bright pupil blobs. Distinguishing the bright pupil blobs from the head motion artifacts in the difference image is a difficult task (Liu et al., 2002).

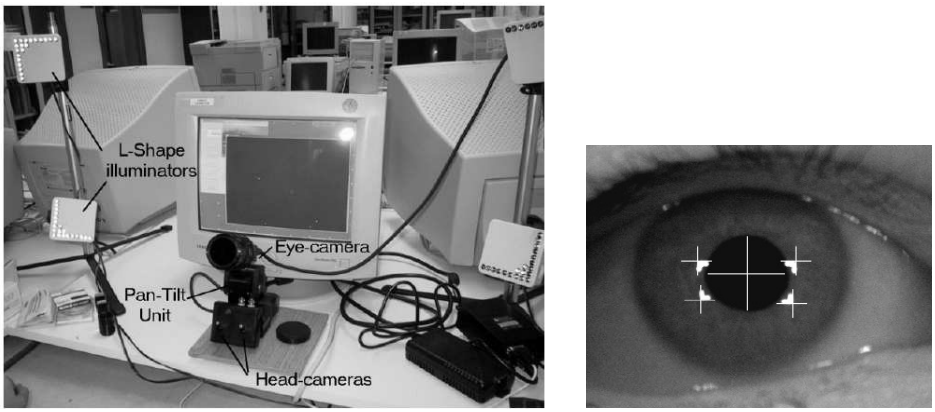
2.3.1.2 Corneal reflection method

A small fraction of light source is reflected off the cornea forming an image of the light source known as a glint (Ji & Yang, 2002). The distinct bright image formed by the glint on the corneal frontal surface can be used as a fixed marker for localizing the eye position, measuring the eye movements, and estimating the gaze direction. The bright intensity produced by glint on the dark iris region can be used to detect eye position using a similar threshold method that is used for the bright pupil detection (Ji & Yang, 2002). Glint also acts as a reasonably fixed marker that can be used to estimate the relative position of the pupil/iris centre in each consecutive frames, which can be used for measuring gaze direction and relative eye movement (Ji & Yang, 2002; Morimoto et al., 2002).

Glint has been used both in head-mounted (Babcock & Pelz, 2004) and remote camera-based eye gaze detection systems (Ebisawa, 1995; Ji & Yang, 2002; Morimoto & Mimica, 2005; Perez et al., 2003). In the remote camera-based gaze direction estimation system developed by Perez et al. (2003), four sets of infrared LEDs were arranged to produce glints with distinct shape, as shown in Figure 2-5. This use of multiple glints with a known shape improved the robustness of their glint based gaze estimation system. Similarly, Yoo and Chung (2005) have also proposed an eye gaze estimation methodology for human-computer interface application based on four light sources on the corners of the computer monitor to produce distinct glint patterns.

Although glints form a good marker for eye detection and gaze direction estimation when they are visible, there are shortcomings that make them unreliable markers. Glints are relatively small and usually require a separate camera with high optical zoom for accurate and robust detection in remote camera-based applications (Ji & Yang, 2002; Perez et al., 2003). The appearance and the visibility of the corneal reflection depend on the face orientation and gaze direction of the eye (Ebisawa & Nurikabe, 2006; Perez et al., 2003). A glint can disappear with a small change in frontal face orientation. Even when the head is stationary, large torsions of an eyeball can shift the cornea far enough for a glint to disappear. The change in curvature between the cornea and sclera can cause distortion to the glint's geometric characteristics and complicate the determination of its image co-ordinates (Perez et al., 2003). For an individual

wearing corrective lenses, the detection of the corneal reflection becomes very difficult due to the glare caused by the reflection of the light source off the corrective lens (Ji & Yang, 2002; Morimoto et al., 2002).



(a) (b)

Figure 2-5. (a) Multiple camera and glint light source based eye gaze detection system. (b) Distinct glints produced by the system (Perez et al., 2003).

2.3.1.3 Summary

The active illumination method capitalizes on the NIR retinal and corneal reflection properties of the eye. These approaches rely on relatively high signal-to-noise ratio. The presence and the brightness of reflections vary on gaze direction, head orientations, and ambient illumination conditions, which are likely to vary in most practical applications. The utilization of the reflective properties of the eye and *a priori* information to detect the eyes and its feature is a novel approach. However, there are other passive facial feature detection approaches that use the *a priori* information of shape, contrast, colour, position, and movement of the facial features and do not rely on a special external light source.

2.3.2 Passive face detection methods

The performance of most passive eye detection system depends on reliable localization of face region of interest (fROI) in an image. Passive fROI localization methods based on colour segmentation (Singh & Papanikolopoulos, 1999; Sirohey & Rosenfeld, 2001; Sirohey et al., 2002; Smith et al., 2003), difference imaging (Morris et al., 2002), and Haar-face detection algorithms were reviewed. The colour segmentation method localizes the fROI in the image by

locating and segmenting the pre-defined facial skin colour. This is likely to have high false positive detection when objects with colour similar to the pre-defined facial skin colour are present in the background image. The difference imaging method localizes the fROI in a video by detecting the head motion. The head motions are detected by subtracting the consecutive frames in a digital video (Morris et al., 2002). In the difference imaging method, any background movement is likely to result in false positive detection. Also, the change of illumination creates motion artifacts in the difference image, which can be falsely identified as head movements by the difference imaging method. Both the skin colour segmentation and difference imaging methods are affected by varying background conditions and are more suitable under constraints environment.

The Haar-face detection algorithm is an attractive choice for fROI localization due to its robustness and accuracy under varying illumination and background conditions (Lienhart et al., 2002). A Haar-face detection application is available with a free open source licence as part of the OpenCV project (OpenCV, 2001). Performance of the Haar-face detection algorithm is comparable to other face detection methods (Viola & Jones, 2001) and is being continually improved with release of new versions of OpenCV. For these reasons, Haar-face detection was used for fROI localization in this project.

2.3.2.1 Face localization using Haar-object detection method

The Haar-object detection algorithm uses a trained-object classifier to localize the object of interest in an image[3] (Lienhart & Maydt, 2002; OpenCV, 2001). For example, the face classifier that defines what a face should look like in an image can be used with the Haar-object detection algorithm systematically search the regions in the image that best fits the face classification. The Haar-object detection algorithm returns the co-ordinates of square regions within the image that are most likely to contain the object of interest.

The Haar-object detection algorithm was introduced by Papageorgiou et al. (1998) who used Haar-like features to form object classifiers. The method was improved by Viola & Jones (2001) to enable it to operate in real-time for face detection with comparable accuracy to other face detection methods. Lienhart et al. (2002) further improved the algorithm by extending the Haar-like features and optimizing the object classifier training algorithm. The Haar-face

[3] Haar-object detection is a generic algorithm that can be used to detect any object defined in the object classifier. In this project the desired object of interest is a face.

detection application in OpenCV uses the cascade of face classifier trained by Lienhart et al. (2002).

Training the object classifiers

The cascade of boosted object classifiers is trained by applying samples of positive and negative images to the modified Adaboost machine learning algorithm (Viola & Jones, 2001). The samples of positive and negative images are the selected images with and without the object of interest, respectively. The images are scaled to same size during training of the object classifier. The cascade of frontal face classifiers used in OpenCV was trained with 5000 positive and 3000 negative sample images (Lienhart et al., 2002). These images were taken from unconstrained environment with varying lighting and background conditions.

The sample images are used to form basic decision-tree object classifiers with subset of Haar-like features. Haar-like features encode the contrast exhibited by an object of interest and their spatial relationship in the image. Examples of a subset of Haar-like features that are used in OpenCV for face detection are shown in Figure 2-6. Haar-like features are an extension of the 2D-Haar wavelet and encode the average intensity between different regions of the image. The Haar-like features are calculated in a similar manner to the coefficients of the Haar wavelet transforms (Papageorgiou et al., 1998), hence the name.

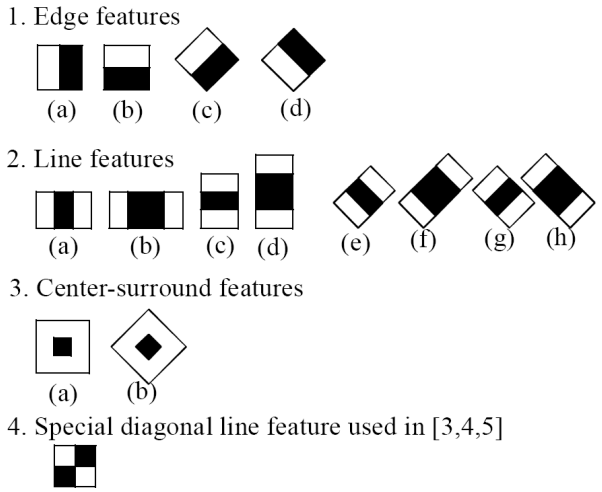


Figure 2-6. Extended Haar-like features used by Lienhart et al. (2002) for face classifier.

The object classifier encodes the shape, position, orientation, and scale of a particular subset of Haar-like features within a region of interest to represent the object in an image. Figure 2-7

shows an example of a way the Haar-like features are used by a face classifier to represent various features of the face (Viola & Jones, 2001). In this example, the horizontal and vertical Haar-like features are used to represent the difference in intensity between the darker regions of the eyes and the lighter regions of the upper cheeks and the bridge of the nose, respectively.

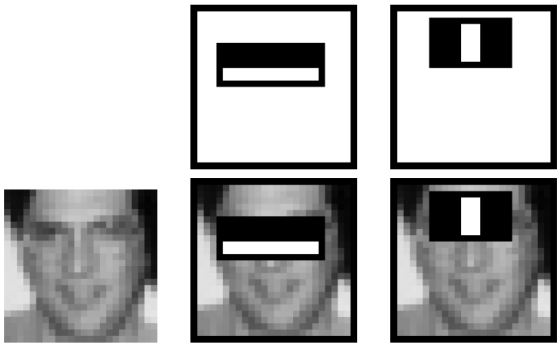


Figure 2-7. Example of Haar-like features used to define different regional average intensity of face (Viola & Jones, 2001).

The object classifiers are used to build more complex boosted classifiers using the various boosting techniques (Lienhart et al., 2002). These boosted classifiers are further combined to be part of the various stages in the cascade structure. The cascade of frontal face classifiers used in the OpenCV is made up of 20 stages. There is a Haar-training application implemented within the OpenCV project that allows adjustment of various parameters to form a cascade of boosted object classifiers (OpenCV, 2001). This training application applies the positive and negative sample images to the machine learning algorithms to form the cascade of boosted object classifiers.

Haar-object detection algorithm

The Haar-object detection method uses the trained cascade of boosted object classifiers to detect the object in any given image. A search window is moved pixel-by-pixel over the entire image to search for the object of interest. The size of the search window is defined by the object classifier. At each pixel, the sub-image within the search window is either rejected at some stage of the cascade of object classifier or accepted if it passes all of the stages. The classifier is designed so that it can be rescaled. The image is scanned several times at different scales of the classifier with increments of 10% of the classifier size to detect objects of unknown sizes. The Haar-face detection application returns either the coordinates of the square regions in the image most likely to contain the face or ‘0’ if no face is found in the image. The

Haar-face detection application developed by Lienhart & Maydt (2002) has a false positive detection of only 24 at a hit rate of 82.3% when applied to a frontal face test set with 510 different frontal faces in 130 grayscale images with 320 x 240 pixel resolution. This hit rate improved with increase in false positive detection of faces.

2.3.3 Passive eye detection methods

The eyes are very noticeable features with many distinct physical and behavioural characteristics. Physical characteristics such as shape, colour, contrast, position, symmetry, and behavioural characteristics such as eyelid and eye movements provide good *a priori* information for passive detection of the eyes and eye features. Most passive eye feature detection systems are based on a top-down model, which first localizes the face region of interest, then the eye region of interest, and finally the eye feature. However, some passive methods can also detect eye features directly in the image. Difference imaging, anthropometric standards, and template matching are some of the passive feature detection methods that have been used to localize the region of interest for the eyes. In addition, computer vision methods employed for passive eye feature detection include edge-detection, active appearance modelling, and projection functions.

2.3.3.1 Difference image

Subtraction of consecutive frames in a digital video produces difference images that can be used for motion detection (Grauman et al., 2001; Kawato & Tetsutani, 2004; Morris et al., 2002). The difference between two consecutive frames due to motion creates pixels with high intensity in the difference image. During the regular spontaneous blinks of a healthy person, the eyelids in both eyes move together. The eyelid movement with blinking creates two high intensity blobs in the difference image as shown in Figure 2-8. The blob detection, optical flow, and erosion methods have been employed in conjunction with the eye's symmetry constraint to extract the position of the blobs created by the blinks in the difference image (Grauman et al., 2001; Kawato & Tetsutani, 2004; Morris et al., 2002). Artifacts in the difference image created by small head motion can be removed by applying erosion methods (Grauman et al., 2001). However, the artifacts created by large head motion swamps the blobs caused by blinking, making it impossible to distinguish them. Although the difference image-

based eye localization method is computationally simple, it is not robust for applications with large head movements.

(a) (b)

Figure 2-8. Difference image with eye blobs caused by blink; (a) before erosion and (b) after erosion to remove small head motion artifacts (Grauman et al., 2001).

2.3.3.2 Anthropometric standards

Proportional facial anthropometric standards define the average proportional distances between facial features across the general population. It can be used to approximate the location of the eyes within the face region in the image (Jin et al., 2004; Lam & Yan, 1996; Zobel et al., 2000). Jin et al. (2004) used the proportional anthropometric standards to localize the eye region with a 91% hit rate in 2375 facial images. Facial anthropometric standards have also been used in graphical face reconstruction and modelling projects (DeCarlo et al., 1998; Kuo et al., 1997). The use of facial anthropometric information within a predefined face region in an image is a computationally simple and efficient method to approximately locate the eye region. However, the localization of the eye region based on the anthropometric standards can tolerate only very small rotations of the head from the frontal face orientation. This eye localization method also relies on accurate detection of the face region of interest.

2.3.3.3 Deformable parametric templates

An eye template can encode various image properties of the eye feature that can be correlated within a given facial image to localize the eyes and the eye features. For example, Tian et al. (2000) encoded the shapes of the eyelids and the iris with two parabolas and a circle, respectively, as shown in Figure 2-9. The parametric values of the parabola and the circle are then varied to measure degree of eye closure and the radius of the iris.

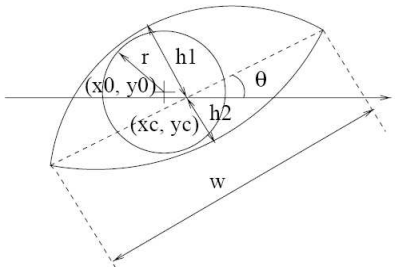


Figure 2-9. Deformable template used by Tian et al. (2000) representing the eyelids with two parabolas and the iris with a circle.

Templates such as two circular iris/pupil rings (D'Orazio, Leo, Cicirelli et al., 2004; Eriksson & Papanikolopoulos, 2001), a single ellipse (J. G. Wang et al., 2003), and semicircular annulus (Sirohey et al., 2002) have also been used to detect iris in an image. In addition to shape, some eye templates also encode the contrast properties of the eye feature. An example of a template that encodes the shape of the iris/pupil and also its contrast is shown in Figure 2-10. This iris/pupil template was applied to the entire image for the exhaustive search of the iris/pupil feature candidates. Overall, this template had the average of 96% successful iris detection when the eyes were open, but performance deteriorated to only 45% iris detection when the eyes were partially or fully closed. Their system operated at 7.5 fps.

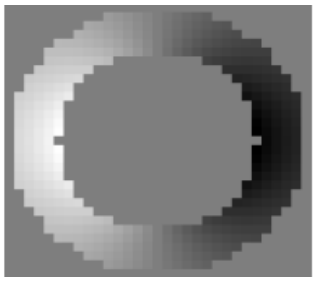


Figure 2-10. Ring template representing the iris and pupil (D'Orazio, Leo, Cicirelli et al., 2004).

The advantage of using deformable templates is that the values of the parameters that encode the features of the eye can be varied allowing the detection of accurate size and orientation of eye features. This knowledge of the exact state of the eye is important for analysing the behavioural characteristics of the eyes. However, the detection of the eye and eye features based on the deformable templates is computationally a very intensive process because all possible parametric values of the template must be correlated with sub-image of a fixed size at every pixel of the entire image.

2.3.3.4 Wavelet templates

Eyes can be detected by processing the image through the wavelet transforms that encode the contrast patterns of the eyes. Sirohey & Rosenfeld (2001) encoded the distinct contrast patterns of the eye region with a Gabor wavelet, as shown in Figure 2-11, to form an eye template. In a similar manner, a Gabor wavelet was used by Zheng et al. (Zheng et al., 2005) to detect the eye corners.

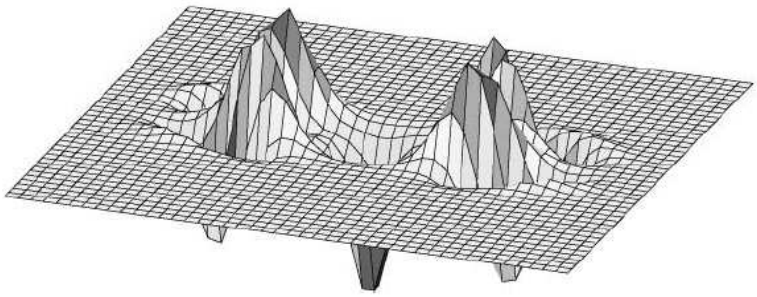


Figure 2-11. Gabor wavelet with emphasized the dark centre iris and surrounding lighter sclera (Sirohey & Rosenfeld, 2001).

The Haar-like features are an extension of the Haar wavelet that is used to construct an eye classifier which encodes the orientation, shape, and size of the contrasts exhibited by an eye (Fasel et al., 2005; Lienhart & Maydt, 2002; Viola & Jones, 2001). The eye classifiers are then used in similar manner to face classifiers to detect the eye region of interest. The Haar-like features based eye detection method is an efficient algorithm for distinguishing the eyes from other facial features in an image. However, the rectangular shape of the Haar-like features can not represent the circular shapes of iris and eyelids in the eye (P. Wang et al., 2005), which makes this method impractical for detecting these eye features.

2.3.3.5 Edge-detection

Eye features can be detected by identifying their edge profiles. Sudden changes in contrast at the edge of the sclera, iris, and eyelids make them good features to detect with an edge-detection method. However, other structures of the eye such as eye folds, eyelashes, wrinkles, and eyebrows produce many undesirable edge candidates which make the task of identifying an edge of any one particular eye feature very difficult. Most of the edge-detection-based eye-

feature extraction systems rely on relatively small search window to minimize false positive edge-detections. Fitting a small search window requires the prior knowledge of the position and size of the eye in the image.

Results with the edge-detection method vary depending on the types of edge operator used and the intensity threshold level set during edge-detection. There are various edge-detection operators, such as Sobel, Roberts, Prewitt, and Canny, each of which define an edge differently and produce varying edge-detection results. Most edge-detection methods require an image intensity threshold that specifies the contrast expected at edge of interest to be set. Different intensity thresholds are likely to produce different edge-detection results. Other factors that can influence the result of the edge-detection-based eye feature extraction system are the glints and the reflection off the glasses.

a **b** **c** **d**

Figure 2-12. Iris edge-detection. (a) Original image, (b) canny edge contour image with many other edge candidates, (c) detection of the two most vertical edges using an edge following technique, (d) circle fitting over the original iris image. (J. G. Wang et al., 2003).

Edge following techniques (Sirohey et al., 2002; J. G. Wang et al., 2003), pre-defined edge contour constraints (Lam & Yan, 1996), and template matching methods (Sirohey et al., 2002) have been applied to identify a particular eye feature's edge in the edge contour image. For example, Figure 2-12 (a-d) illustrates the edge following and circle fitting methods used by Wang et al. (2003) to extract the edge of an iris.

Shortcomings, such as the variation in the result depending on the selection of the edge operator and the intensity threshold and also the difficulty of distinguishing the edge of the particular eye feature from many undesirable edge candidates, make the application of the edge-detection method unsuitable for eye feature extraction.

2.3.3.6 Active appearance model

Ishikawa et al. (2004) developed a global 2D face model with an active appearance modelling (AAM) method for a remote camera based gaze estimation system. Their AAM defines the shape and appearance of the face with 2D triangulated mesh. Once the 2D face model is accurately fitted onto the subject's face in the image, the positions of the eyelids and eye corners are derived from the model. These positions are tracked over time to detect the eyelid movements and gaze directions. The template matching method with dark-disk iris template is applied within the eye region to detect the position of the centre of the iris. The movement of the iris centre relative to the locally static positions of the eye corners are tracked to estimate the gaze direction. The gaze estimation system developed by Ishikawa et al. (2004) based on AAM face model and iris template matching is shown in Figure 2-13.

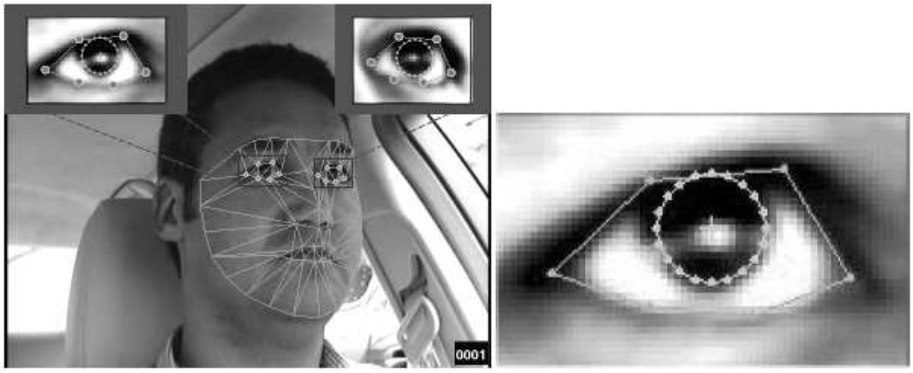


Figure 2-13. The eye gaze detection system developed by Ishikawa et al. (2004) based on the AAM of the face and iris template matching methods.

2.3.3.7 Image projection functions

Image projection functions plot the intensity distribution of an image into a 2D plot which is easy to analyse. Three main types of image projection functions that have been employed for eye and eye feature detection: integral projection function, variance projection function, and general projection function (Zhou & Geng, 2004). An integral projection in the vertical direction is calculated by averaging the intensity across all columns in each row of the image. Horizontal integral projection is calculated in a similar manner by calculating the mean intensity across all rows in each column of the image. In the variance projection function, the variance of the columns and rows are calculated instead of the mean. The general projection function is the weighted sum of the integral and variance projection functions.

A distinct change in intensity from light to dark to light at the eye region in the face and also the intensity contrast between the eye features within the eye region can be detected by analysing the projection function of the image. The eye region corresponds to the highest intensity peak in the vertical integral projection in a facial image, as shown in Figure 2-14. This cue is used to approximate the vertical position of the eye within the face region (Eriksson & Papanikolopoulos, 2001; Haisong Gu & Ji, 2004; Singh & Papanikolopoulos, 1999). The change in image intensity between eye features such as eyelid-iris, iris-sclera, and sclera-eye corner can also be detected by analysing the projection function within the eye region of interest (Zhou & Geng, 2004).

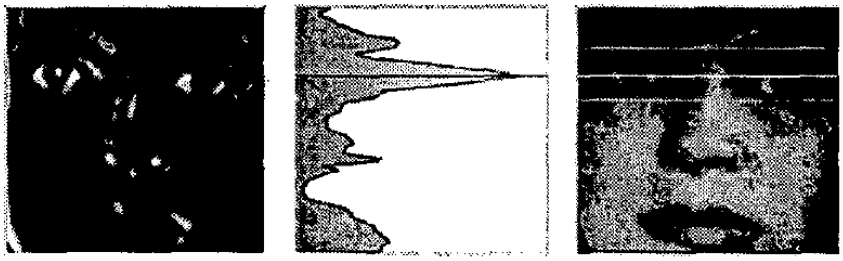


Figure 2-14. The vertical integral project of the face to extract the vertical eye position (Singh & Papanikolopoulos, 1999).

When various image projection functions were applied to three different facial databases, Zhou & Geng (2004) found that all had a eye detection rate of over 94%. They also found that the variance and the general projection functions were more effective for eyelid and eye corner detection than the integral projection function (Zhou & Geng, 2004). In addition, the integral projection function was better at detecting eye feature in oriental faces, while variance projection function performed better with occidental faces (Zhou & Geng, 2004). In contrast, Zheng et al. (Zheng et al., 2005) reported that the variance and general projection functions have less reliability for the general population. The image projection functions are computationally very efficient but are only reliable for eye feature extraction under small eye regions of interest. The presence of any excess accessory eye structures like eyebrows will cause the performance of the image projection functions to deteriorate. The image projection functions are also affected by variation in illumination and reflection off corrective lenses.

In general, passive eye feature detection methods perform best with a top-down model, where the face region of interest is first defined followed by the eyes regions of interest, before the fine features of the eye such as eyelids, eye corners, and iris are detected. The top-down

approach provides confirmation at every step of the process that the right region in the image is being searched, which improves the reliability of the system. Passive methods are reliable for detecting an eye feature because they use visual cues from multiple surrounding eye features. In this study, an active NIR lighting source was employed for invariant illumination and passive methods were used for face, eye, and eye feature detection.

2.4 OpenCV computer vision library

OpenCV is an open source computer vision project that aims to provide a development platform for computer vision algorithms with collection of libraries and applications (OpenCV, 2001). It provides a cross-platform middle-to-high level API with hundreds of image processing and computer vision C functions and C++ classes. It is a free software library with both non-commercial and commercial licenses (OpenCV, 2001). The project currently has a wide range of functions ranging from basic functions for image manipulation to more advanced functions like Kalman filters. In addition, these functions and classes have also been used to develop popular computer vision algorithms. Relevant algorithms in OpenCV that were investigated are the Lucus-Kanade tracking algorithm (Bouguet, 1999), the CAMSHIFT algorithm (Bradski, 1998), and the Haar-object detection algorithm (Lienhart et al., 2002). The Haar-face detection application developed as part of the sample applications in OpenCV has been directly integrated into this project. OpenCV also provides I/O libraries for easy video data acquisition and manipulation from multiple camera inputs. In this project, the video frames from the camera were acquired using the “highgui.dll” I/O libraries in OpenCV. OpenCV was a very useful tool for rapid prototyping and algorithm implementation during the current project.

2.5 Commercial video-based drowsiness detection systems

Growing awareness of fatigue and drowsiness contributing to a substantial proportion of motor vehicle and work-related accidents has created a demand for commercial drowsiness detection system. The development of such system has been of particular interest for automobile companies which have funded many drowsiness detection research projects (Desai & Haque, 2006; Ji & Bebis, 1999; Ji & Yang, 2002; Lal & Craig, 2001; Ueno et al., 1994). Some of the independent computer vision technology companies that have developed systems for facial

metrics measurement that are particularly marketed for drowsiness detection: Seeing Machines (Australia), SmartEye (Sweden), and Attention Technologies (USA).

FaceLAB (Figure 2-15) developed by Seeing Machines, Australia, is a real-time system with two 60 Hz small form factor digital FireWire cameras that measure head pose, eyelid movement, gaze direction, and other facial metrics. The FaceLAB system tracks each eye independently and if the subject is close enough, it also tracks the size of the pupil. FaceLAB can operate in varying illumination conditions with the subjects close to or several metres away from the cameras. The system can track up to 6 degrees of freedom (DOF) head movement and 2 DOF gaze direction within 1° accuracy. Several automobile and medical companies, as well as research groups, have incorporated FaceLAB in their applications. FaceLAB is aimed at a wide range of industrial and research applications such as medical, transport, biometric security, human-machine interface, robotics, and child psychology research studies. FaceLAB is quoted at NZ$35,000.

Figure 2-15. Dual camera-based FaceLab system developed by Seeing Machines (www.seeingmachines.com/index.htm).

AntiSleep developed by SmartEye, Sweden, is a facial feature detection system marketed towards automotive driver fatigue and attention detection (AntiSleep, 2005). AntiSleep measures head position, head orientation, gaze direction, and eyelid closure using a single 60-Hz VGA-resolution CMOS camera and two infrared-flash illuminators as shown in Figure 2-16(a). In this system, an infrared filter is also used to minimize interference from the outdoor light. A fully automatic initialization procedure of the AntiSleep system detects the generic and special facial features and maps them into a generic 3D head model. The head model is then adapted to the driver in real-time. This system can operate with various types of eyeglasses and also handles reflections off the eyeglasses.

AntiSleep measures the head position within 10 mm accuracy and head orientation within 4° accuracy. The gaze direction is estimated with the 10° accuracy and the eyelid closure (vertical distance between the eyelids) with 2 mm accuracy. AntiSleep is an embedded DSP based product. An example of its PC interface is shown in Figure 2-16(b). AntiSleep system costs 25,000 Euros (approximately NZ$ 46,000).

(a) (b)

Figure 2-16. (a) Single camera and infrared LED based AntiSleep drowsiness detection product and (b) its PC interface, developed by SmartEye (www.smarteye.se).

The DD850, shown in Figure 2-17, is an NIR retinal-reflection-based commercial system developed by Attention Technologies, U.S.A. DD850 is based on the methods developed by the "Copilot" research project in Robotic Institute at Carnegie Mellon University, USA (Grace et al., 1998). This system uses bright pupil occlusion detection to measure the PERCLOS parameter for evaluating driver alertness level. It is a dashboard mountable device with embedded DSP to perform the image processing task. The DD850 warns the drowsy driver with an audible alarm and provides the LED-based visible feedback about the alertness level to the driver.

Figure 2-17. DD850, the NIR retinal reflection based drowsiness detection system developed by Attention Technology (www.attentiontechnology.com).

Chapter 3 System design and initial experiments

This project aimed to develop algorithms for automatically detecting the face and eye features using video data, which, in turn, can be used for measuring the facial metrics associated with drowsiness and microsleep. Section 3.1 in this chapter proposes the facial-feature detection algorithms that must be developed and how they can be used to measure the facial metrics associated with drowsiness and microsleep.

The video-based system would initially be used in conjunction with other sleep metrics (section 1.2) in lapse research studies by researchers in CNRP. The CNRP ultimately aims to develop and commercialize a device (or devices) that can use a combination of video-based metrics, EEG, EOG, and motor kinematic metrics to (1) warn users of their state of deep drowsiness and the likelihood of imminent lapses and/or (2) detect the onset of lapses and provide wake-up alarms. To make the video-based system practical in real-world applications, several operational requirements were identified and are presented section 3.2. Then, section 3.3 and 3.4 presents the design considerations for video data acquisition to address some of the operational requirements.

Initial investigations of several potential computer-vision methods identified in Chapter 2 were carried out and are presented in section 3.5. In section 3.6, a brief overview of the processes of system development carried out in this project is presented.

3.1 Proposed system

Figure 3-1 shows a flow chart of the face and eye feature detection algorithm that must be ideally implemented to estimate the facial metrics associated with signs of drowsiness and microsleep. The components above the dashed line in Figure 3-1 represent the computer vision passive feature detection methods required to measure three main video-based facial metrics (section 1.8) which can, in turn, be used to measure the components below the dashed line that represent the behavioural metrics of drowsiness and microsleeps (sections 1.7).

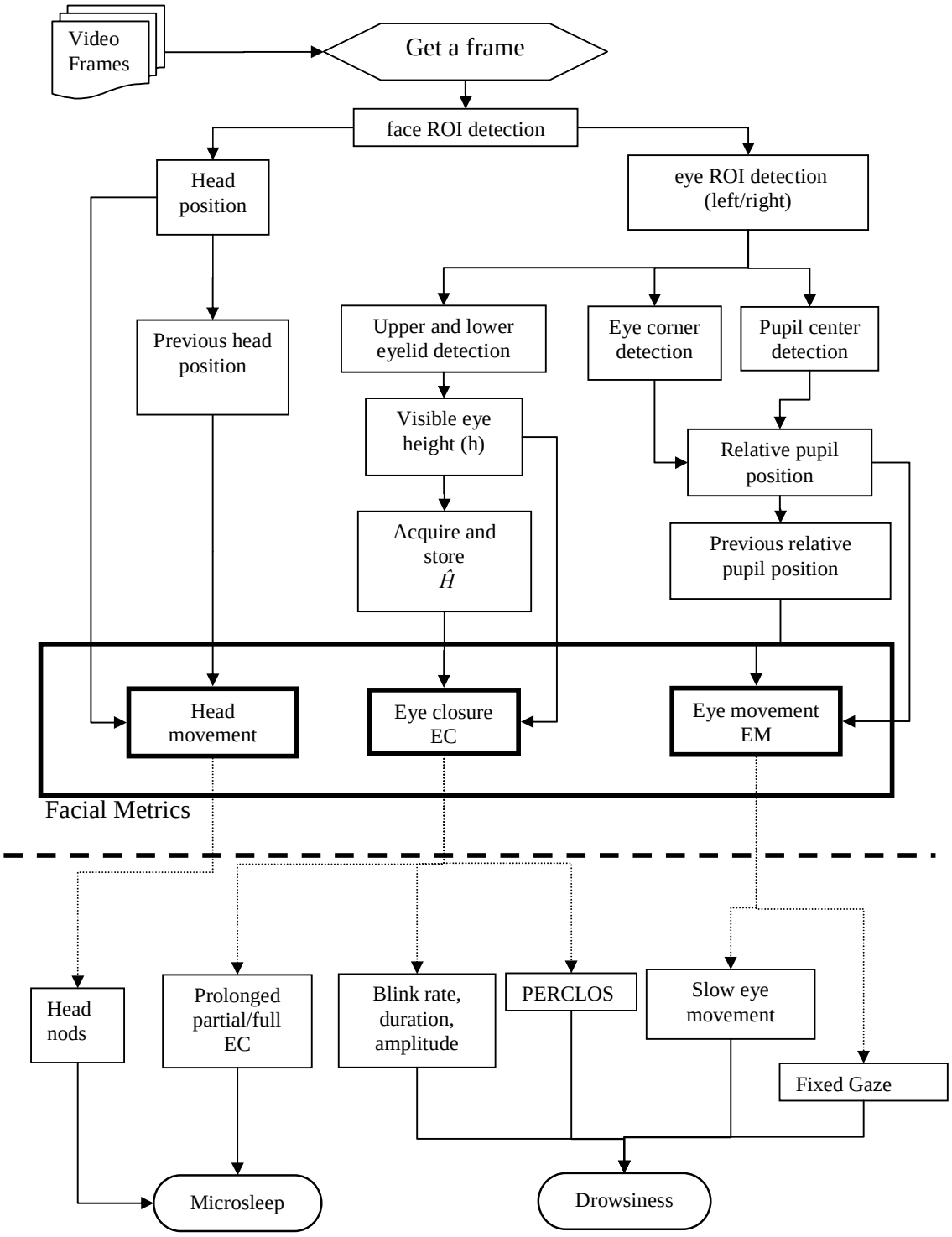


Figure 3-1. Flow chart of an ideal video-based microsleep and drowsiness detection algorithm.

This project aims to develop and implement the passive algorithms to detect the face and the parts of the eyes which will be used for measuring the facial metrics. Although NIR illumination was used for active lighting of facial scene (as explained later in section 3.4), NIR retinal and corneal reflection properties of eyes were considered to be unreliable from a review of relevant literature (Chapter 6) and preliminary experiments (presented later in section 0). Hence, the retinal and corneal reflections off the eyes were not utilized for detecting the position of eyes in this project. Instead, passive feature detection methods were used for localizing the face and eye features.

The passive feature detection approach usually involves narrowing down the search region of interest before localizing the feature of interest. To measure the facial metrics, the face region of interest (fROI) must be initially estimated within the image. Estimation of the fROI in an image reduces the search area for the eyes within the image. Furthermore, the estimated position of the fROI in consecutive frames of a video can also be used to measure the 2D translational head movement by subtracting the position of the face in current frame from its position in previous frame.

Localization of the eye region of interest (eROI) in an image improves the performance of the passive eye feature detection methods by reducing the likelihood of detecting false positive results and also reducing the computational load. Once the eROI is defined, the system must detect the upper and lower eyelids to measure the visible eye height. To estimate the degree of eye closure, the system must initially acquire and store the mean visible eye height from the frames with fully open eyes. Once the mean visible eye height is acquired, the degree of eye closure within each frame can be calculated as the ratio of visible eye height in the current frame to the mean visible eye heights of fully open eyes. Within the eROI, the eye corners and the centre of pupil are the other features of the eyes that must be detected to measure the relative position of the centre of pupil. By subtracting the relative position of the centre of pupil in the current frame from its relative position in the preceding frame, the eye movement can be measured. These three video-based facial metrics can then be used to identify various behavioural signs of drowsiness and microsleeps.

3.2 Operational requirements

A video-based alertness monitoring system must be practical and end-user compliant to be useful in real-world applications. It should be designed so that it can be incorporated into a

wide range of unconstrained real-world environments such as in the instrument panel of a car or an aeroplane, in front of an air-traffic controller, and even inside a hospital's operating theatre. In most applications, the video-based alertness monitoring system must be able to operate without distracting the end-user or compromising their natural operating environment, while maintaining its accuracy and robustness in varying ambient conditions. The basic system design requirements considered important in an automated video-based alertness monitoring system are that it should:

1. be non-intrusive to allow the end-user to move freely and to operate in their natural operating environment without additional constraints or distraction,
2. work under unconstrained ambient environment such as varying backgrounds and illumination levels, including darkness,
3. work for people across the population, regardless of differences in facial structure and colours,
4. work in the presence of corrective lenses,
5. operate in real-time.

Design considerations to meet some of the operational requirements of the video-based system are described in the next two sections.

3.3 Non-intrusive remote camera-based system

To make the video-based system non-intrusive to the user, the face and eye feature-detection algorithms in this project were developed around video data acquired from a remote camera placed at a fixed distances from the user. A remote-camera-based system allows free head rotations and translation, at least to certain extent, without the face being lost from the camera's field of view. Figure 3-2 shows an image of a subject acquired with a remote camera placed at approximately 60 cm in front of the subject. In this image the subject's face is centred in the camera's field of view with sufficient background visible on either side of the face to capture any lateral head movement. One face width on either side is incorporated in the field of view. Unlike a head-mounted system, the remote-camera based system can be incorporated discretely away from the user's sight and without impairing their vision or distracting them from their task.

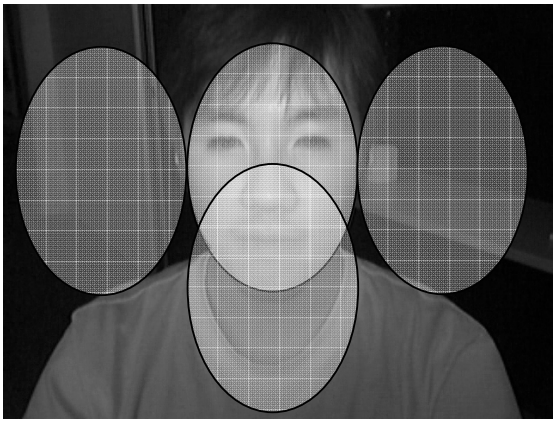


Figure 3-2. An example of an image acquired from the remote camera placed at approximately 60 cm away from the subject with one face width space on either side of the face to allow head movement.

A Logitech QuickCam 4000 camera, designed for webcam application, was used for digital video data acquisition in this project. The camera utilizes a glass lens and a Sharp LZ24BP CCD sensor which acquires digital images with 24-bit RGB colour and 640 x 480 pixels VGA resolution. The camera streams the digital images at 30 fps through a USB 2.0 interface.

The LZ24BP CCD sensor is a ¼-type progressive-scan sensor with RGB primary colour mosaic filters and is sensitive to both the visible and NIR spectrum. Since the camera is intended to operate in the visible spectrum, it has an optical highpass filter in front of its lens to block the IR spectrum. This IR-block optical highpass filter was removed so that the camera could operate under NIR illumination for reasons explained in section 3.4.

The glass lens in the QuickCam 4000 camera allows acquisition of better quality images than the plastic lenses common on cameras for webcam use. A built-in automatic white balancing feature of the QuickCam 4000 camera compensates for varying ambient illumination levels by controlling the electronic exposure rate to reduce variation in image intensity, making the camera robust under varying illumination. However, the camera also allows the electronic exposure rate to be manually adjusted if required.

For the video-based alertness monitoring system to be useful, the camera must sample the images fast enough to capture the eyelid closure during a spontaneous blink while a person is alert. During a typical spontaneous blink, the eyelids are more than 75% closed for approximately 66 ms corresponding to a bandwidth of approximately 15 Hz (Evinger et al., 1991). To meet the Nyquist sampling requirement, the camera must therefore sample the images at a rate of 30 Hz or greater to detect closed eyelids during a typical spontaneous blink.

The sampling rate of 30 fps through the USB 2.0 interface by the QuickCam 4000 was considered to be adequate to operate the video-based alertness monitoring system in real-time because the eyelid movements observed during the drowsy periods are relatively much slower than spontaneous blinks while alert. The bandwidth for data transfer from the camera to a computer is another factor that must be considered for real-time operation. Although the 480 Mbit/s bandwidth of the USB 2.0 interface is sufficient for video data transfer in real-time, the host-centric nature of the USB interface requires hand-shaking overhead between the host computer and the webcam during which video frames can be lost. The peer-to-peer nature of the FireWire interface does not require hand shaking and is better suited than the USB 2.0 interface for video data transfer in real time applications.

Although real-time operation is one of the requirements for the video-based system to be useful, the development of robust and accurate algorithms was considered to be of higher priority during this project development. Hence, the development of the algorithms was based on post-processing of images from the recorded videos. For this reason, the video data acquisition at high frame rate through the FireWire interface was not critical during the project development. However, a camera with a higher frame rate and FireWire interface should be considered to operate the system in real-time in future.

Webcams are designed for general purpose video data streaming applications such as video conferencing over the internet and are not ideal for machine vision applications. Cameras that are designed specifically for machine vision applications with higher image quality and operating specifications than webcams are also available in the market. However, the image quality of the high-end low-cost QuickCam 4000 camera was considered sufficient for initial development of the computer vision facial feature detection algorithm. In addition, the performance of the face and eye feature detection algorithms developed based on the images acquired from the low-cost webcam is likely to improve when applied to the higher quality images acquired from a camera specifically designed for machine vision application.

3.4 Visible light insensitive system

Infrared emitting diodes (IRED) and an optical infrared lowpass filter (OILF) that blocks out visible light were utilized to make the video–based system relatively insensitive to a wide range of visible lighting conditions. Under low visible lighting or dark conditions, the subject’s face is illuminated with the NIR illumination emitted from the IREDs. NIR illumination is

ideal for illuminating the scene under dark conditions because, unlike the bright glare from a visible light source, NIR is invisible to the eyes and will not distract the subject from their task. A set of 6 OPE5587 IREDs with peak emission at 880 nm NIR wavelength and bandwidth of ±45 nm was used for illuminating the facial scene in this project. The optical highpass filter in the QuickCam 4000 camera which blocked the infrared radiation was removed to allow the CCD sensor to detect NIR illumination. Figure 3-3 shows an image of lit IREDs as acquired by the QuickCam 4000 camera without the IR-block optical highpass filter.

Figure 3-3. An image of the lit IREDs acquired by the QuickCam 4000 with the IR-block filter removed.

Under ambient visible lighting condition, the system was made insensitive to visible light by blocking out the visible spectra and acquiring the images purely under NIR illumination. The visible light was blocked out by placing a #87 gelatin-base OILF in front of the CCD sensor in the camera. The #87 OILF is opaque under visible light but allows more than 85% transmission of the electromagnetic radiation with wavelength above 875 nm. Since the IRED illuminates the subject's face with the NIR illumination and OILF allows the CCD sensor to detect only the infrared illumination. Hence, a video-based system can operate effectively regardless of the visible lighting condition. Figure 3-4 illustrates the setup of the IRED-based NIR illumination source, the IR sensitive camera, and the OILF used for video data acquisition in this project.

However, if the NIR illumination-based system is used under outdoor conditions, the level of intensity in the NIR images will be influenced by the infrared component in sunlight. Unfortunately, the setup of the IRED and the OLIF cannot avoid the variation in image intensity due to changes in level of sunlight exposure under outdoor conditions such as driving the vehicle from the shaded area to area exposed to direct sunlight and vice-a-versa. Under varying NIR illumination levels, the built-in white-balancing feature of the camera should compensate the varying image intensity to some extent by controlling the level of exposure.

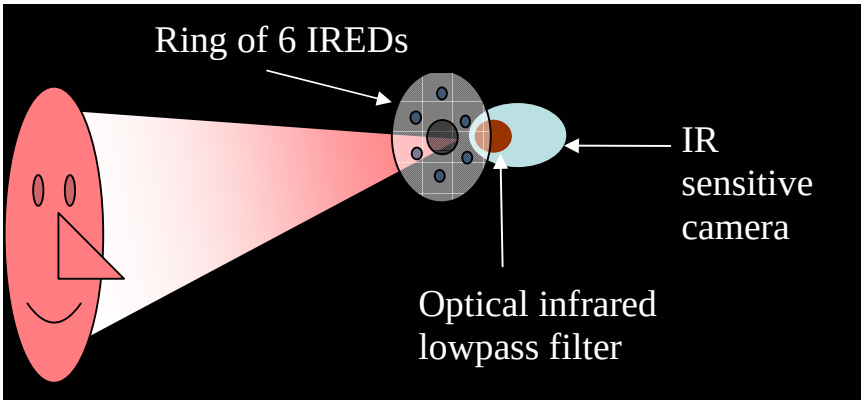


Figure 3-4. The setup of the IR sensitive camera, the IREDs and OILF to make the system invariant to visible lighting condition.

3.5 Initial trials for eye and eye feature localization

At the initial stage of this project, experimental trials of potential methods for localizing eye position and centre of iris/pupil were carried out to investigate their feasibility under the desired operational environments (section 3.2). Eye localization methods based on both the active and passive approaches as discussed in literature review (Chapter 2) and novel ideas were investigated. The corneal reflection and retinal reflection based active eye localization methods were investigated due to the ability to directly detect eyes without first detecting face within an image. Initial trials of the passive eye localization methods were carried out using the software and face classifiers for the Haar face detection algorithm that are readily available in the OpenCV library. The passive eye localization methods investigated are blink detection, iris template matching, and edge-detection methods. These initial experimental trials provided an important exercise that highlighted their limitations under the project's operational requirements.

3.5.1 Eye localization from corneal reflection

As discussed in section 2.3.1.2, corneal reflection or glint has been used by different researchers as a good marker of an eye in remote camera-based systems. It was considered appropriate to investigate its reliability to locate eyes, as the NIR source is used for

illumination of the scene in this project. Figure 3-5 show a NIR-image with corneal reflection from an experiment.

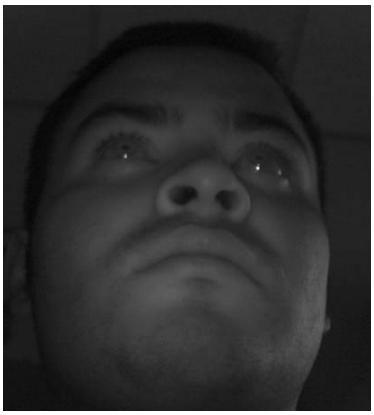


Figure 3-5. NIR image showing corneal reflections of NIR emitting diode.

Although corneal reflection can form a distinct marker of an eye, its presence and visibility in an image was unreliable. As found by others (Ebisawa & Nurikabe, 2006; Perez et al., 2003), the corneal reflection depends on the face orientation and the eye-gaze direction. In Figure 3-5, the corneal reflections disappeared with the slightest upward head pitch. Presence of any other NIR source also made the marked corneal reflection less distinguishable. Similarly, during the day the presence of NIR component of the sun made corneal reflection very hard to distinguish. In addition, the corneal reflection method was found to be impractical in subjects wearing glasses because it was often overwhelmed by the reflections of NIR off the glasses.

3.5.2 Pupil localization from NIR retinal reflection

A method for using retinal reflection of NIR illumination (bright pupil effect) to localize eyes in an image is discussed in section 2.3.1.1. In an experimental trial, the bright pupil effect was mostly visible in completely dark conditions because the larger diameter of the pupil due to iris dilation under dark lighting conditions allowing more amount of NIR illumination reflected off the retina to escape. Figure 3-6 shows an example of bright pupils produced during the initial experimental trial.

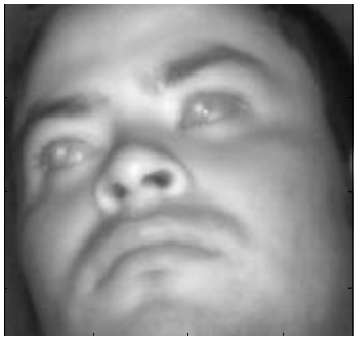


Figure 3-6. Example of bright pupils that was generated during initial trials.

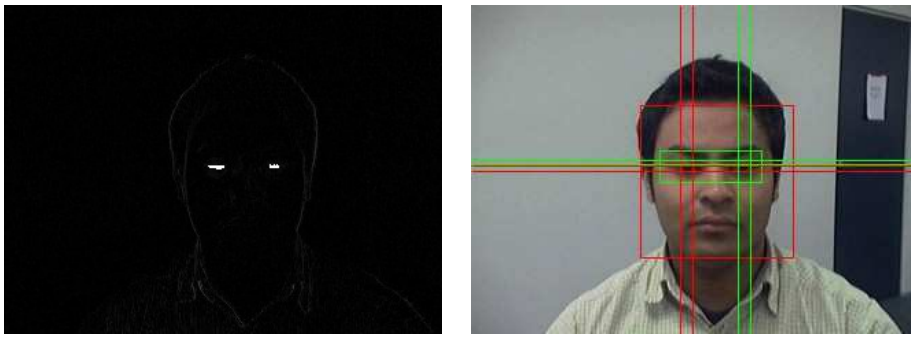
However, the constriction of the iris in ambient indoor lighting condition reduces the diameter of the pupil which makes it difficult to produce bright pupil. Increasing the intensity of NIR illumination did improve the reliability of presence of bright pupil. Subjects also reported a small increase in the facial temperature and tired eyes. Hence, increasing the NIR illumination was considered unsafe. In addition, low resolution of the image acquired from the web-cam made it harder to see the bright pupil effect during the ambient lighting condition. A combination of low resolution image and small pupil size in ambient lighting condition also made it difficult to differentiate retinal reflection from corneal reflection. Finally, as in the corneal reflection-based eye detection method, reflection of NIR illumination off spectacles made it difficult to distinguish bright pupil in subject who wore glasses.

3.5.3 Eye localization from difference images of blinks

As discussed in section 2.3.3.1, eye blinks can be used to localize eye positions in a video with frontal facial images by detecting blobs formed in difference images that are derived by subtracting consecutive frames in the video. A short sequence of videos with eye blinks was collected and a simple blob detection function was developed and applied to the videos to investigate the feasibility of the method to localize eye positions.

Figure 3-7 (a) shows a difference image where blobs are formed as the subject's eyes blink and Figure 3-7 (b) shows the localized positions of eyes in the current frame based on detection of the blobs. To improve the method performance and computational efficiency, an eROI within the difference image was defined by scaling the fROI localized by Haar-face detection algorithm (section 2.3.2.1). The eROI was scaled based on proportional facial anthropomorphic

knowledge (section 2.3.3). Within the eROI, the blobs created due to blinks were verified based on their size, threshold intensity, and distance between two adjacent blobs.



(a) (b)

Figure 3-7. (a) an example of a difference image showing blobs formed during a blink; (b) Localization of the eyes based on blob detection in the difference image in (a).

The blob detection in a difference image was a simple and an effective method for localizing eye positions when the subject remained stationary. However, artifacts created by large head motion swamped the blobs created by eye blink, as shown in Figure 3-8, making it impossible to distinguish them. Although the difference image-based eye localization method is computationally simple, it is not robust for applications which must accommodate head movements.

Figure 3-8. Difference image derived by subtracting consecutive frames captured during head movement the motion artifacts due to head movement in the eROI makes it difficult to detect the eyes.

Since head movement relative to camera is inevitable in the remote camera-based system adapted in this project, the blink detection method was impractical for eye localization.

However, due to its simplicity and computational efficiency it can perhaps be used in conjunction with other appropriate methods to provide second reference positions of the eyes, particularly since in most task intensive applications such as driving a vehicle the head is likely to be stationary over short periods of time.

3.5.4 Localization of iris with a disk template matching

The diagram in Figure 3-9 illustrates the template matching-based method initially trialled for detecting centre of iris and its radius.

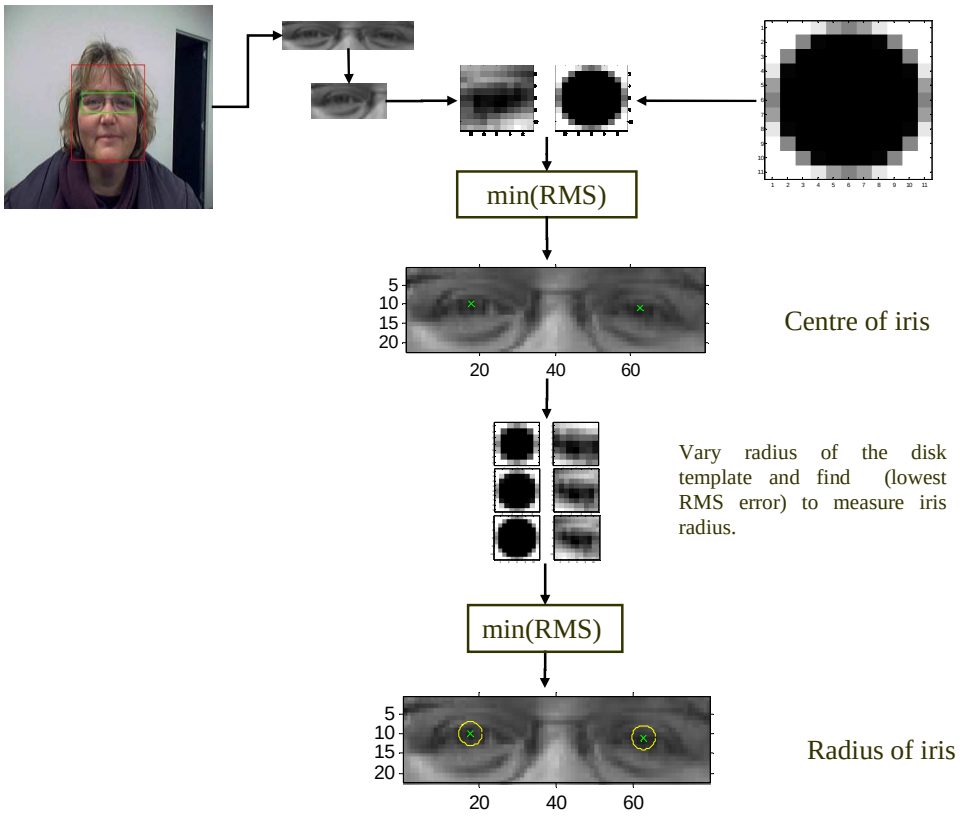


Figure 3-9. Diagram illustrating template matching-based method for detecting centre and radius of an iris.

Firstly, the eROI is equally divided into the left and the right regions. In each eROI, the root mean square (RMS) error is calculated between a simple dark disk image template representing an iris and sub-images of same size as the template centred at each pixel in the eROI. The position of a pixel with the lowest RMS value is selected as the centre of the iris. To determine the radius of the iris, the radius of the disk template is varied while calculating the RMS error

between the template and the sub-image of same size as the disk template centred at the estimated centre of iris. The radius of the template which gives the least RMS difference is considered the best estimate of the radius of the iris.

Visual inspection showed that the position of the iris/pupil was detected within a close proximity (within the dark circular region of the iris) of the true centre of iris in 76% of 192 frames analysed. However, the eROI had to be reduced to tightly fit the eyes to get the best result. Figure 3-10 shows two examples of eROI with correctly estimated centre of iris and radius whereas Figure 3-11 shows two examples of where the iris template matching methods failed in estimation of iris centre (left eROI) and radius (right eROI).

Figure 3-10. Examples of correct estimation of iris centre and radius.

Figure 3-11. Example of false detection of centre of iris (left) and incorrect estimation of iris radius (right).

Although, this initial trial to localize the position and estimate the radius of the iris using the template matching method was not satisfactorily robust, it gave promising results. However, template matching method was computationally intensive and slow. The localization of the iris/pupil position is important for measuring eye movement and gaze direction. The methods for measurement of these metrics were not developed in duration of this thesis. However, the encouraging result from this initial trial led to adopting the template matching method for detecting the centre of eye to measure eye closure as explained later in section 5.6.

3.5.5 Edge-detection

Various edge-detection operators were applied to a selected set of eROIs. The output of each edge operator was visually analysed for ability to outline the edges of eye features. Figure 3-12

shows an example of edge-detection results obtained by applying Sobel, Perwitt, Roberts, and Canny edge operators (Forsyth & Ponce, 2003) to an eROI.

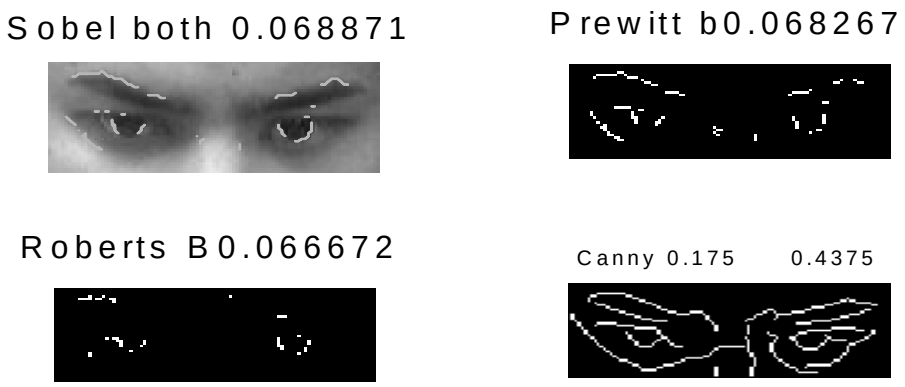


Figure 3-12. Example of edge-detection in an eROI using various edge operators and their corresponding intensity threshold derived automatically as implemented in edge() function MATLAB™.

As discussed in section 2.3.3.2, the following shortcomings of the edge-detection method make it unattractive for eye localization within an image:

- Requirement of small ROI for good results,
- Difficulty distinguishing edges of desirable eye features,
- It is difficult to identify correct intensity threshold and dynamic threshold algorithms are not perfect.

3.5.6 Conclusion

The foregoing experimental trials were carried out in the initial stage of the project to gain experience with different methods for localization eye and iris/pupil positions. Each of the initially trialled methods for localizing eye position had shortcomings that discouraged further investigations. Particularly, the active localization of pupil using the corneal and retinal reflection of NIR showed worse than expected results. However, the template matching based iris localization method gave encouraging initial result. In addition, using proportional facial anthropomorphic information was a reliable method for deriving an eROI from fROI in frontal face images. Hence, from these experimental trials, it was decided to pursue the passive face and eye feature detection approach.

3.6 Project overview

After the initial trials, the development and evaluation process for the passive facial feature detection system was initiated. As shown in Figure 3-13, this process was performed in three main stages:

1. Collection of reference video data and annotation of selected frames.
2. Development of face and eye feature detection algorithms.
3. Automated performance evaluation of developed algorithms, with results used to improve performance of the algorithms.

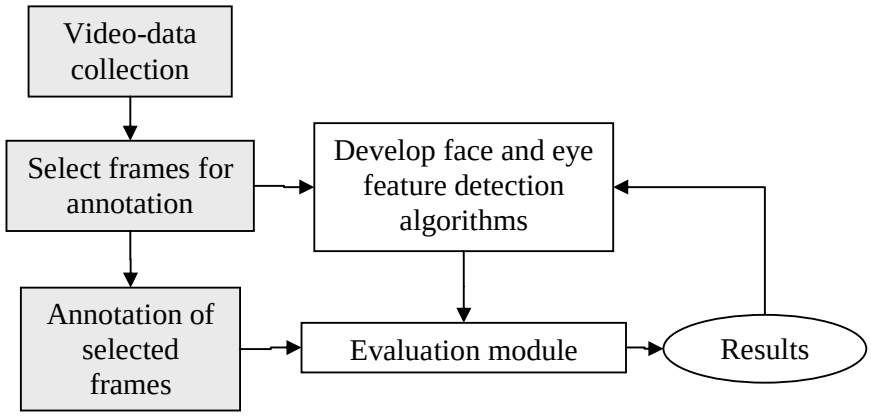


Figure 3-13. Block diagram of the main tasks and the data path in the project.

These three stages of project development are presented in rest of the chapters of this report. In Chapter 4 the process of reference video data collection and manual annotation of selected reference frames are presented. Then Chapter 5 presents the detail description of the developed face and relevant eye features detection methods and their evaluation using the selected annotated frames.

Chapter 4 Reference data collection and annotation

The collection of sample video data is important for determining and analysing the characteristics of images under varying ambient conditions and traits of facial features. Information from this can be used in the development of robust and accurate feature detection algorithms. Furthermore, it is important for the sample video data to be annotated with true reference information on the features of interest for quantitative performance evaluation of the developed algorithms. This, in turn, will help identify the strengths and weaknesses of the algorithms. Generic facial image databases, such as BioID (BioID, 2001) and FERET (Phillips et al., 2005) with annotated positions of the face and eyes, are available in public domain for performance evaluation of facial feature detection algorithms (Wu & Zhou, 2003; Zhou & Geng, 2004). However, these databases contain neither images under NIR lighting condition nor subjects exhibiting signs of drowsiness and microsleeps. Hence, to develop algorithms for video-based alertness monitoring system that can operate under NIR illumination condition, reference videos of nine subjects imitating various actions, including signs of drowsiness and microsleep, were recorded under four different recording conditions. From the videos of each subject, sets of 66 frames were selected and manually annotated with the position of the face, the eyes, and the relevant features of the eyes to form the reference data for qualitative and quantitative performance evaluation of algorithms developed. This chapter presents the methodology for the reference video data collection and annotation.

4.1 Video data collection

The video data were collected on the basis of an experimental design which defined the experimental setup, criteria for selecting subjects based on variation in traits of facial features, actions to be performed by the subjects, and video recording conditions.

4.1.1 Experimental setups

Figure 4-1 shows the experimental setup used for video data collection. Subjects were seated on a chair in front of a desk with a computer monitor where they could see what was being recorded. The camera and the six IREDs arranged in a 7 cm diameter ring concentric to the camera were fixed on the top-centre of the monitor. Once seated, subjects were asked to adjust

their seating so that the front legs of the chair were horizontally aligned to a distance marker on the floor, giving the eye-camera distance of 60 cm. This distance was chosen from trials to achieve optimal (see section 3.3) field of view of one head width on either side of the face. Once the subject was seated correctly, the field of view of the camera was adjusted so that the subjects' face appeared frontal and approximately in the centre in the image plane. Also the focal length of the camera was manually adjusted to focus on the face.

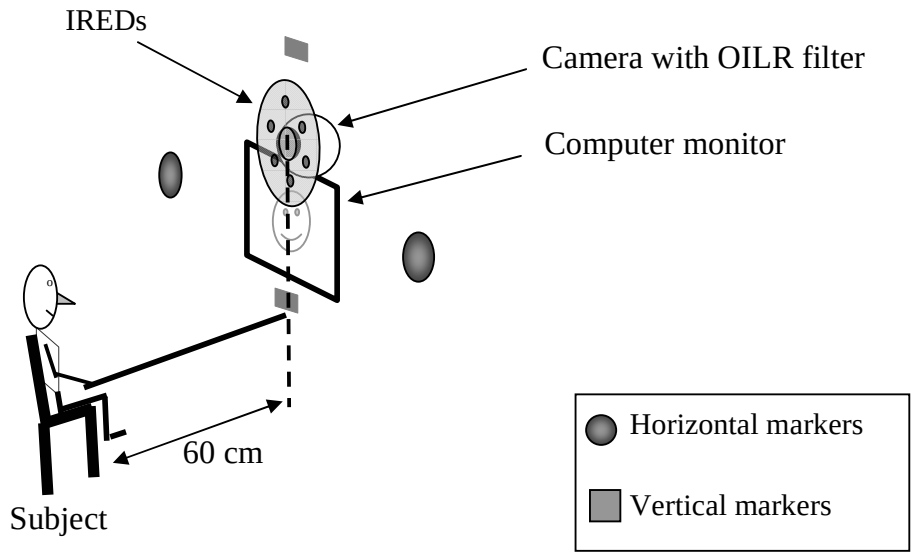


Figure 4-1. Schematic of experimental setup for video data collection.

To achieve consistent gaze direction between subjects and to define instruction for the horizontal and vertical gaze actions during the video recordings, two horizontal and two vertical stationary markers were respectively placed in front of them as shown in Figure 4-1. The horizontal markers were placed in the same plane and height as the fixed camera at approximately 30˚ adjacent angle horizontally in front of the subject. Similarly the vertical markers were placed in the same plane as the camera at approximately 45˚ adjacent angle vertically in front of the subject.

4.1.1.1 Infrared exposure safety

During reference video data collection it was important to limit the infrared exposure to the subject's eyes below the ICNIRP's (International Commission on Non-Ionizing Radiation Protection) maximum infrared safe exposure limit of 10 mW/cm^2 irradiance for exposure longer then 16 min (Matthes, 2000; Sliney, 2000). Under the NIR illumination video

recordings, an exposure of 0.15 mW/cm^2 NIR irradiance was estimated from the six OPE5587 IREDs at 60 cm. This level of irradiance was calculated from the radiant intensity and half angle information obtained from the datasheet of the OPE5587 IRED. In support of this calculation, an NIR exposure of approximately 0.2 mW/cm^2 irradiance was measured by the J16 Tektronix Digital Photometer/Radiometer with the J6502 infrared probe at 60 cm from the six IREDs. Hence, the infrared exposure to the subject was safe and well below the maximum safe infrared exposure limit set by ICNIRP.

4.1.1.2 Data acquisition hardware

The video data were acquired with QuickCam 4000 camera connected to a PC via a USB 2.0 interface. A program 'CamSerialCtrl.exe' was developed for controlling the camera setting, acquiring and storing the video data on a hard drive, and sending a switching signal via a RS232 interface to a microcontroller-based IRED control circuitry. The software was implemented in C++ and used the OpenCV project's "cvCam" I/O library for acquiring the camera properties and video frames from the camera. Images acquired by the camera were 640 x 480 pixel arrays with 24-bit colour resolution at a frame rate of 30 fps. The built-in auto-white balancing feature of the camera was turned on to automatically control the exposure rate of the camera. The incoming video data stream was compressed to DivX 3.1 MPEG 4, Low-Motion video format at bit-rate of approximately 955 Kbit/s and stored as AVI video file format in a hard drive.

Figure 4-2 shows the circuit diagram of the microcontroller-based switching circuitry for the IRED. In response to switching command from the user via an assigned key press on the keyboard, the 'CamSerialCtrl.exe' program sends ASCII data via the RS232 interface to control the switching of the IREDs. The Atmel ATMEGA163 microcontroller was programmed in embedded C to receive, interpret, and execute commands sent to it from the PC via the RS232 interface. On receiving the appropriate command, the microcontroller pulls its I/O port (to which the switching circuitry for the IREDs was connected), high or low to turn the IREDs on or off respectively.

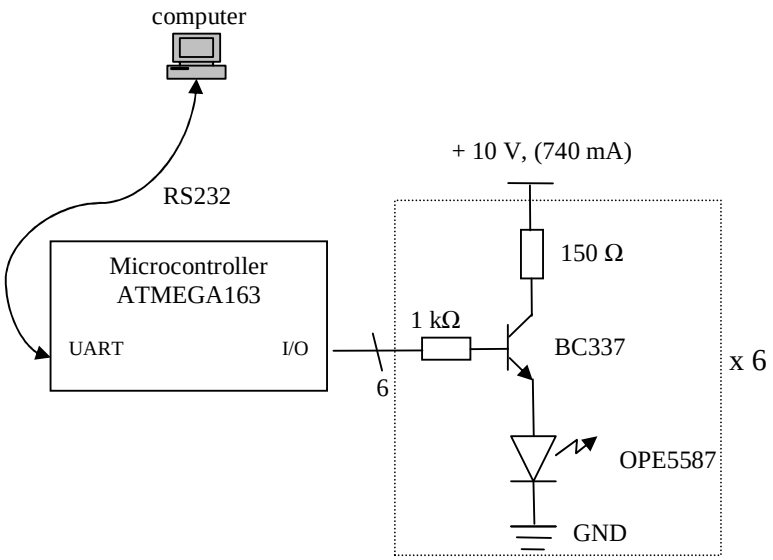
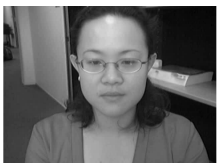
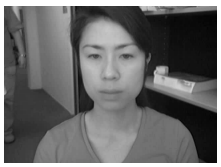


Figure 4-2. The circuit diagram of the microcontroller based IRED switching circuitry.

4.1.2 Subject selection process

For reference video data collection, nine subjects (five male and four female) were selected and instructed to perform various eye and head movement tasks relevant to the state of drowsiness and distraction. The subjects were selected so as to gain a relatively wide representation of variation in facial features across gender, race, and wearing of glasses because the video-based system must work for people across population and in presence or absence of glasses. Five subjects were of Asian origin (3 with single eye folds, 2 with double eye folds) and four were of European origin. Four of the nine subjects wore glasses. Table 4-1 summarizes the relevant facial feature of the nine subjects selected for reference video data collection. Variations in facial features between races can affect the performance of feature detection algorithms. For example, the integral projection function is more effective at detecting eyes in occidental faces then in oriental faces and vice versa with the variance projection function (Zhou & Geng, 2004). Zhou & Geng attributed variation in performance of the algorithms to shadow of the nose and eyeholes between races. Some of the other examples of variation in eye features in the population are colour of the iris, colour of eyebrows and eyelashes, and types of eyelid folds.

Table 4-1. Summary of relevant information on subjects selected for reference video data collection.

Subject no.	Gender	Glasses	Race	Hair	Iris/ eye folds	Photos
1	Male	No	Asian	Dark	Dark/ double	
2	Female	No	European	Light	Dark/ double	
3	Male	No	Asian	Dark	Dark/ double	
4	Male	No	Asian	Dark	Dark/ single	
5	Female	Yes	Asian	Dark	Dark/ single	
6	Female	No	Asian	Dark	Dark/ single	

7	Female	Yes	European	Light	Light/ double	
8	Male	Yes	European	Light	Dark/ double	
9	Male	Yes	European	Dark	Light/ double	

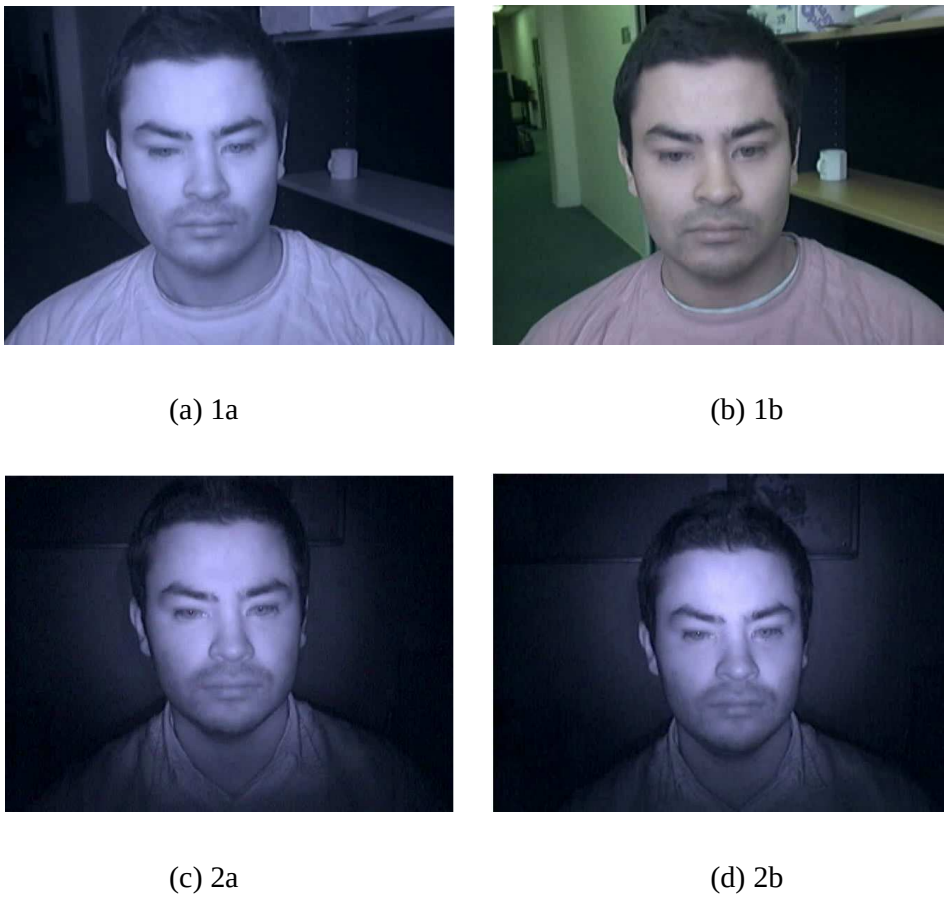
4.1.3 Recording conditions

The videos of all nine subjects performing the same set of actions were recorded under four different recording conditions:

1. Day-time condition with subject seated inside a room with ambient florescent lighting and next to the windows exposed to sunlight and
    a. with NIR illumination and the OILF placed in front of the camera
    b. without artificial NIR illumination or OILF
2. Night-time condition with subject seated inside a dark room without any ambient lighting or windows and
    a. with NIR illumination and the OILF placed in front of the camera
    b. with NIR illumination but without the OILF

Figure 4-3 shows an image from each of the corresponding videos recorded in the four recording conditions. In videos recorded under day-light conditions, as shown in Figure 4-3(a) and (b), the scene was lit by the NIR illumination from the IREDs and the infrared component in sunlight. Under recording condition 1b, both the visible colour spectra and the infrared spectra were visible in the image, as show in Figure 4-3(b), because no optical filters were placed in front of the camera. Under recording condition 2a, the image, as shown in Figure

4-3(c), is formed purely from NIR illumination emitted by the IREDs. The only other source of illumination in the dark room conditions was the LCD monitor whose brightness was set to minimum level and which does not emit NIR illumination[4]. The IREDs were not strong enough to illuminate the entire background of the scene. Hence, in the images under condition 2a, only the face of the subject is clearly visible. When the OILF was removed in condition 2b, the background became slightly more visible in the image, as shown in Figure 4-3(d). This was due to the background being illuminated by the LCD monitor and the CCD sensor now being able to detect this visible illumination reflected from background.



(a) 1a (b) 1b

(c) 2a (d) 2b

Figure 4-3. Images from the videos recorded under four respective recording conditions: (a) day time with the NIR illumination and OILF, (b) day time without the NIR illumination and OILF, (c) dark room with the NIR illumination and OILF, (d) dark room with the NIR illumination but without OILF.

[4] The LCD monitor used during video data collection did not emit any infrared spectra. This was confirmed by observing an image of the monitor captured by placing the OILF, which blocked visible light, in front of the camera. Although the monitor was turned on and the programs running on the monitor were visible, it appeared blank and turned off in the IR image.

4.1.4 Recorded events

In each video, subject was instructed to carry out a set of actions that were designed to imitate eyelid movements, eye movements, gaze directions, and head movements that the video-based alertness monitoring system must detect to identify subject drowsiness, microsleep, and distraction. The eyes and the head movements were performed in two subsequent sessions:

Session 1: Eye movements while keeping the head stationary and facing straight a head

a. Spontaneous blink
b. Slow droopy eye closure followed by full eye closure for at least one second
c. Rolling the eyeball horizontally from side to side to gaze towards the stationary horizontal markers
d. Rolling the eyeball vertically up and down to gaze towards the stationary vertical markers

Session 2: Head movements

e. Head nod by simultaneously drooping both the eyes and the head followed by jerky head movement back to upright head position.
f. Head nod by drooping the head while keeping the eyes open at all times
g. Head nod by drooping the head while keeping the eyes closed at all times
h. Slowly pan the head from side to side to gaze towards the stationary horizontal markers
i. Slowly pitch the head up and down to gaze towards the stationary vertical markers

The subjects were instructed to perform each action three times sequentially. Additional natural spontaneous movements, such as blinks and saccades, were not removed from the collected video data.

4.2 Annotations

To evaluate performance of eye feature detection algorithms, a set of frames from videos of each subject was selected. In each of these frames, the position of the features of interest in the face and the eye regions were manually annotated to form a reference database for performance evaluation of developed algorithms.

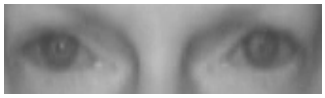
4.2.1 Frame selection process

Manually annotating the features of interest in each frame of all four recorded videos would have been a very time-consuming and rigorous task and not feasible or necessary in this project. Hence, only a few important frames were selected for manual annotation. The frames were selected based on the specific conditions the video-based alertness monitoring system must operate in and variation in eye features under different eye movements that it must detect. For each of the nine subjects, 2 sets of 33 frames from each of the two videos recorded under conditions 1a and 2a (i.e., videos recorded under NIR illumination and OILF in ambient light and dark conditions) were selected for annotation.

To limit the number of frames selected for annotation, only the frames with the subject in frontal face orientation were selected. Ideally, an alertness monitoring system would need to be able to detect facial features under all head orientations, but as an initial step, it was considered sufficient to have the feature-detection algorithms developed operating robustly with frontal face orientation as this is, by far, the prominent facial orientation in which the alertness monitoring system will have to detect eye features in majority of applications. For example, when a driver or pilot is drowsy and likely to have a microsleep, they are most likely to be facing and looking straight ahead. In fact, if they do not have a frontal facial orientation in the image from the camera, it is not unreasonable to assume that they are alert but distracted.

Eyelid closure and eye movement are important facial metrics the alertness monitoring system must be able to quantify accurately and reliably. To determine the ability of an eyelid closure measurement algorithm to identify different levels of eye closure, frames with five degrees of eye closures were selected. Similarly, to evaluate the performance of the developed eye movement measurement algorithm, frames with six eye positions relative to head while the subject is facing straight a head were selected. In this report, since only the frontal face frames were selected for evaluation, the ‘gaze’ direction is interchangeably used with the eye position relative to frontal face image. To determine the reliability of the algorithms, three frames for each of the 11 categories of eye closure and gaze were selected. Table 4-2 lists the 11 categories of eye conditions selected for annotation.

Table 4-2. Example image of 11 categories of eyelid closure and eye gaze frames selected for annotation.

Frame selection categories	Example
1. Eyes wide open	
2. Eyes 3/4 open	
3. Eyes 1/2 open	
4. Eyes 1/4 open	
5. Eyes closed	
6. Left gaze	
7. Extended left gaze	
8. Right gaze	
9. Extended right gaze	
10. Extended upward gaze	
11. Extended downward gaze	

The frame selection process involved observing each frame in the video and subjectively deciding if the frame represented one of the 11 frame selection categories. The sequence of frames in the video was navigated backwards and forwards to detect and select the frames that best represented the selection categories. For example, when a frame with fully closed eyes was identified, the next few frames were analysed to make sure that the eyes started to reopen and not close further. The frames were selected from random position in the video sequence. A frame selected for a particular category was labelled with a number 1-11 to indicate the frame selection category and the frame number in the video sequence was recorded.

In summary, total of 594 frames were selected for manual annotation, comprising of 66 frames for each of the 9 subjects, and 33 frames in each of the light and dark conditions. The 33 frames comprised of 3 frames for each of the 11 eyelid closure and gaze direction categories.

4.2.2 Feature annotations

The true position and the dimensions of facial features, particularly regarding the eyes, were manually annotated in all the selected frames. The annotation process involved either visually marking a point or fitting an appropriate shape to a feature of interest in an image. A point marker represented by the intersection point of an 'x' was used to indicate the x and y coordinates of the position of a feature in an image. In contrast, an ellipse or rectangle was fitted to a feature in order to extract its position coordinates and dimensions. Coordinates and dimensions could be measured at the sub-pixel level. Figure 4-4 shows one of the selected images with annotation of face and eye features.

Nose-bridge, eyebrows, and visible parts of the eyes were annotated on the face. The nose-bridge and eyebrows are not critical features for alertness monitoring but provide approximate boundary points for the eyes within a face. The coordinates of the nose-bridge and eyebrows were each annotated with two points. The nose-bridge was marked with upper and lower points approximately in a centre point between the eyes and tip of the nose. Each of the left and the right eyebrows were approximately marked with inner and outer points. The inner point of an eyebrow was marked at the edge of the darkest medial region of the eyebrow and the outer point was marked at the lateral local maximum of the eyebrow curve.

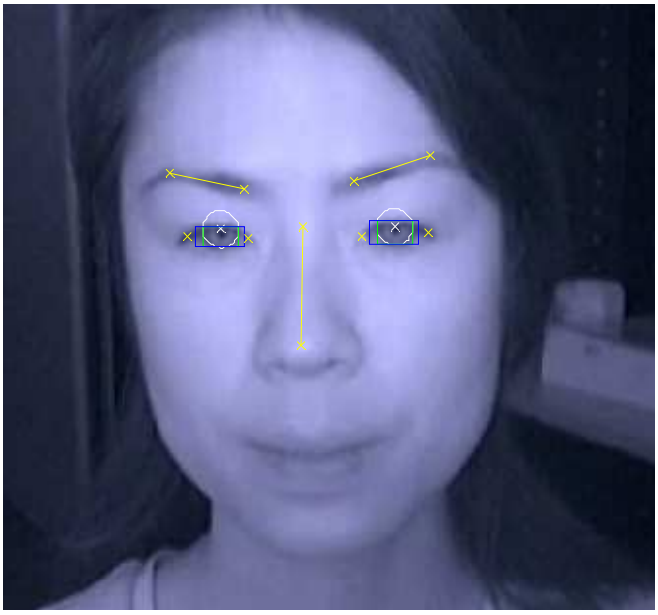


Figure 4-4. Annotated features in one of the selected frames.

Eye features were annotated with more care and precision then the nose-bridge and eyebrows. Figure 4-5 shows the markers used for annotating the iris, eyelids, and eye corners. An iris was marked with a point in the centre of the pupil, an ellipse was fitted to the edge of the iris, and a rectangle was used to mark the area of the visible iris. When the eyes were open, the visible part of the eyeball was marked with a rectangle, whose horizontal edges aligned with the apex of the upper and lower eyelids. Hence, the height of the visible eye's rectangle indicates the height of visible eye for performance evaluation of the eye-closure measurement algorithm. The vertical edges of the visible eye rectangle were horizontally aligned to the horizontal extremes of the visible eyeball.

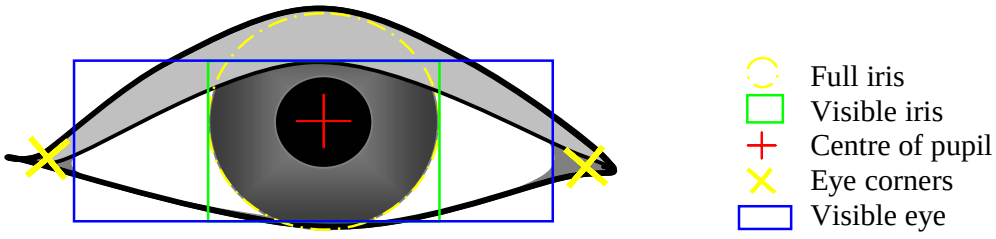


Figure 4-5. Schematic of annotated eye with annotation markers.

Obviously, the eyeball was not visible in images with fully closed eyelids. Hence, for evaluation purposes, the annotation markers of the iris were ignored in the frames labelled to have fully-closed eyes and the height of the visible eye was set to zero. Also, the bottom horizontal edge of the rectangular visible eye was used for marking the vertical image coordinate of the position where the upper and lower eyelids met, as illustrated in Figure 4-6.

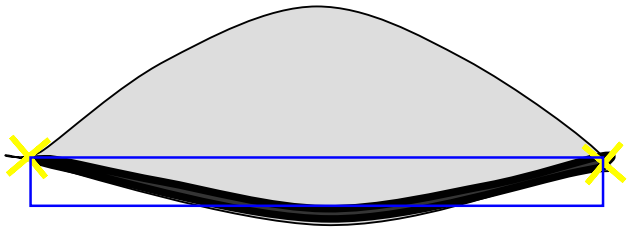


Figure 4-6. Annotation of a closed eye.

The inner and outer corners of the eyes were marked with two points in the facial image. It was found difficult to consistently annotate the eye-corner points in all frames for a particular subject because the eye-corners are ill-defined, for example, due to eye lashes occluding the eye-corners. Hence, the inner corner of an eye was annotated as a point within the lacrimal caruncle, medial commissure, and medial canthus and the outer corner was annotated as a point within the lateral commissure and lateral canthus.

4.2.3 Annotation data structure

The annotated data for each set of selected frames were stored in a data structure defined by the subject number and the video-recording condition. In addition, the path to the directory in the hard drive of the recorded video was also stored so that the selected frames could be acquired from the video data. Table 4-3 lists an example of an annotation data structure for subject number 4 (‘P4’) under the ambient daylight (L1), NIR illumination (I1), and OILF (F1) video-recording condition.

For each of the 33 selected video frames, the frame numbers representing the position of the frames on the video sequence, the corresponding number associated with the eye condition in each frame, and the position coordinates and dimensions of each annotated facial feature were

recorded in the data structure. The annotated data were further sorted into left eye, right eye, and nose-bridge data structure. For features annotated with a point marker, such as iris centre, nose-bridge, and eyebrows, x and y coordinates of the point were stored. Annotation of the iris edge with an ellipse comprised of x and y image coordinates for the centre and the value for the major and the minor radii of the ellipse. The annotated data for features marked with a rectangle, such as visible iris and visible eye, were represented in the data structure with x and y image coordinates of the top-left corner, width, and height of the rectangles.

Table 4-3. Annotation data structure format.

Subject number: 'P4'
Recording condition: 'L1I1F1'
Video file path: 'Y:\ Lapse Detection Project\Data\RefVidDB\ \P1\RefVidData_1_L1I1F1.avi'
Eye condition: [1x33] e.g., [2, 5, 4, 2, 1, 1, 5, 3, …]
Frames number: [1x33] e.g., [47, 69, 71, 73, 129, 131, 887, 1245, …]
Annotation data: [1x33]

Nose bridge: [x_{upper}, x_{lower}, y_{upper}, y_{lower}]
Left eye:
Right eye:

Iris centre: [x, y]
Full iris: [x, y, major, minor]
Visible iris: [x_{left}, y_{top}, width, height]
Eye corners: [x_{inner}, x_{outer}, y_{inner}, y_{outer}]
Visible eye: [x_{left}, y_{top}, width, height]
Eyebrows: [x_{outer}, x_{inner}, y_{outer}, y_{inner}]

4.2.4 General discussion on annotation process

The selected frames were carefully annotated by the author. However despite this care, there will be a certain level of human bias and error involved with manual process. In addition, the consistent precision annotation was very challenging because of the relatively low image resolution around the eye region of only 50 x 200 pixels on average, as shown in Figure 4-7, and further reduction in image quality under low illumination.

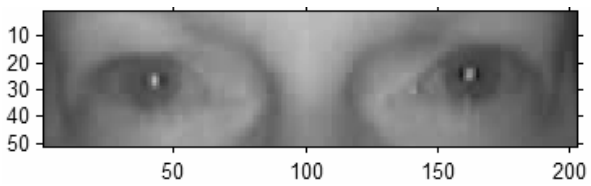


Figure 4-7. Average pixel resolution of 200 x 50 pixels around the eye region.

Ideally, the frames would have been annotated by multiple expert-makers to take inter-expert marker variability into account. However, the time consuming process of feature annotation deemed annotation by multiple expert-markers unfeasible in this project. Nevertheless, the ability of an automated computer-vision feature detection system to perform comparable to that of a single human expert's feature recognition ability should be satisfactory at this stage.

Chapter 5 Facial feature detection

5.1 Overview

This chapter describes methods implemented in this project for detecting face region of interest (fROI), eye region of interest (eROI), centre of eye (COE) position, and eyelids positions in an image. In addition, detected eyelids are used for measuring eye closure. Figure 5-1 shows a top-down approach flow chart of the facial feature detection and the eye closure measurement methods developed in this project. Initially, each colour image in video data was converted into 8-bit grayscale resolution image. Then, the Haar-face detection algorithm was used to localize a fROI within the grayscale image. Once the fROI was localized, predefined proportional anthropometric scaling constants were defined and used to scale the fROI to derive eROI.

Within an eROI, the position of an eye was estimated by applying a template matching method to detect the COE position. An eye template that encoded the average contrast in intensity and shape of the eyes in the reference database was formed and used for COE detection. Once the position of COE was estimated, the size of the eROI was further optimized for eyelid detection. The y-coordinates of apex of upper eyelid (UEL_y) and lower eyelid (LEL_y) were detected by analysing the gradient of vertical integral projection of the optimized eROI. The estimated y-coordinate of the COE was used as the position to start the search of the eyelids because it is vertically positioned between the upper and lower eyelids. The detected eyelid positions were used for measuring the visible eye height (h) of an eye, which, in turn, was used with the reference height ($\hat{H}$) of the eye when it is fully open to calculate the fractional eye closure (EC). The sections in this chapter give detailed descriptions of each of the top-down facial feature detection methods.

5.2 Converting RGB images to grayscale

Since all of the feature detection methods used in this project primarily use contrast in image intensity and do not rely on colour cues, the input RGB colour images are converted into 8-bit grayscale images. The grayscale value for each pixel is computed by taking the root mean square of the corresponding R, G, and B components. The grayscale images make the computation for feature detection algorithm simpler and less intensive.

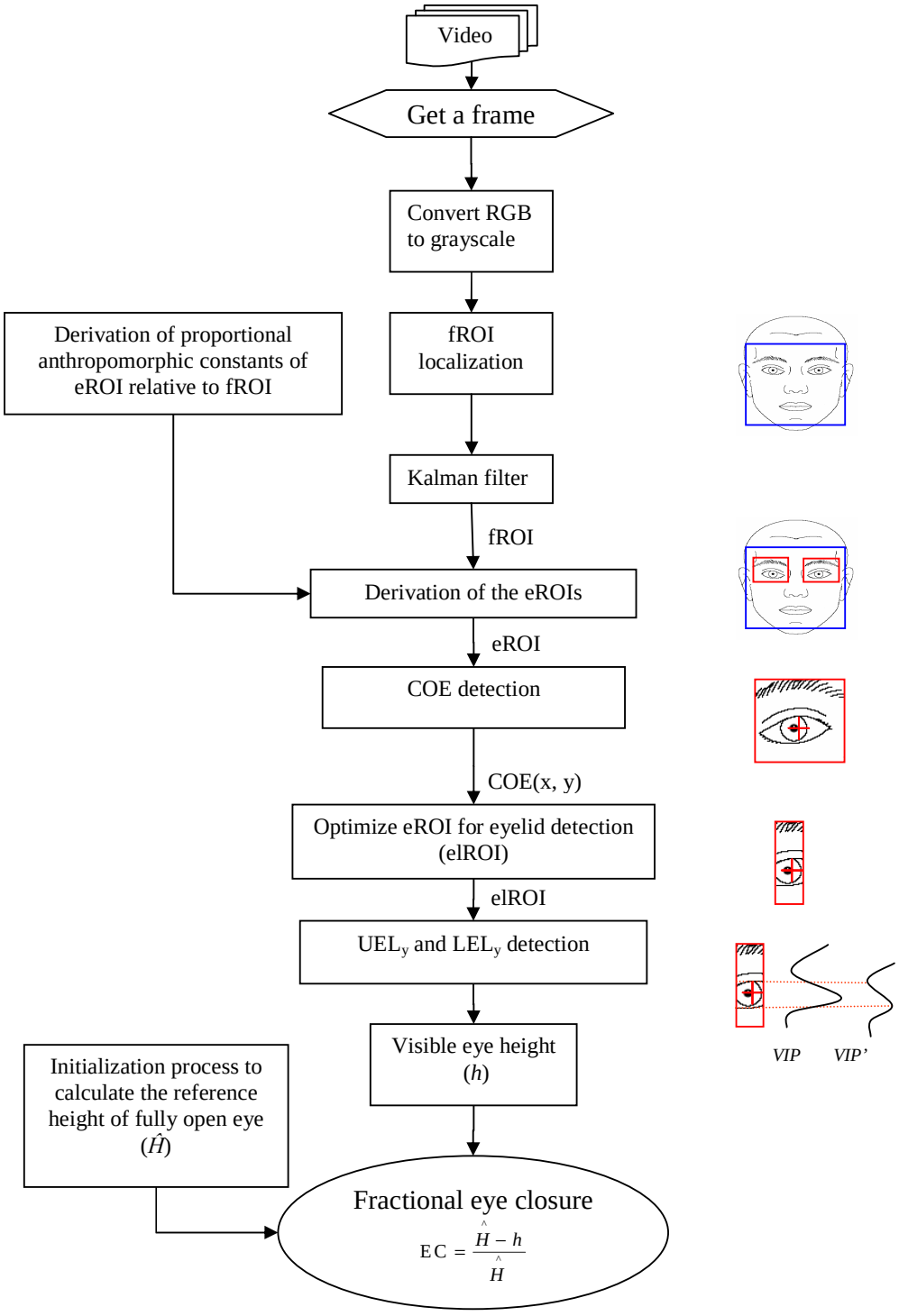


Figure 5-1. Flow chart of the methods developed for measuring eye closure.

5.3 Detection of face region of interest

The Haar-object detection algorithm is a generic algorithm that searches for an object specified by input object classifier (see section 2.3.2.1). In this project, the Haar-object detection algorithm and a face classifier which are implemented in software by Lienhart & Maydt (2002) and readily available in the public domain as part of the OpenCV libraries was used for detecting a face region of interest (fROI) in an image. Figure 5-2 shows an example of a fROI detected by the Haar-object detection algorithm and the frontal-face classifier in one of the grayscale images.

Figure 5-2. An example of a face region of interest (fROI) detection with Haar-face detection algorithm implemented in OpenCV. The fROI is a square defined by a top-left corner coordinate (fROI(x_{min}, y_{min})), a size parameter.

The Haar-object detection algorithm is implemented as C-language functions and the face classifiers are stored in XML file format in OpenCV. The function prototype of the algorithm is shown below. The function acquires seven input arguments and returns square fROI candidates in the image. Each output fROI coordinate is defined by top-left corner coordinate (fROI(x_{min}, y_{min})) and a size parameter.

```
cvHaarDetectObjects(image, cascade, storage, flags, min_size, scale_factor,
                    min_neighbours)
```

The *image* to be analysed is passed in as an input pointer argument to the Haar-face detection function and all possible output fROI candidates are stored in the *storage* input argument. The *cascade* input argument defines the object classifier to be used. In this project, a particular frontal-face classifier "haarcascade_frontalface_alt.xml" available in OpenCV was used for classifying a face. The frontal face classifier was used because the images in this project were acquired with frontal face.

Adjustment of the other four input arguments in the Haar-object detection function allows a trade-off in the speed and the accuracy of its output. This project was initiated with aim to achieve the best performance of the developed methods and establish a baseline for future developments. Hence, the values for each of these four input arguments were conservatively set to acquire high accuracy in fROI detection while compromising processing speed.

Setting the *flags* input argument causes the function to apply the Canny edge detector to the image and rejects regions that does not contain a predefined number of edges expected to be detect for a face. This reduces the image area that the algorithm has to scan and makes the algorithm faster. However, this heuristic approach compromises the accuracy of the Haar-face detection algorithm by scanning less image area and possibly rejecting area where the face actually exists. Hence, the input argument *flags* was left unset in this project.

The *min_size* input argument of the function specifies a smallest expected size of fROI. Initially, the algorithm scans the input image with a face classifier that is scaled to the size specified by *min_size*. Specifying larger *min_size* will increases the processing speed of the algorithm as larger region of image will be scanned. However, large *min_size* will constrain the detection of smaller face in the image. In the images acquired in this project, the faces of all subjects approximately occupy one sixth of the 480 x 640 pixel image. Hence, the minimum size was conservatively set to 80 x 80 pixels.

The Haar-object detection algorithm detects the variation in size of a face by scanning an image multiple times with different scales of face classifier. In each subsequent scan, the initial *min_size* of the face classifier is enlarged by a proportion specified by the *scale-factor* input argument. For example, when the *scale-factor* is set to 1.1 as in this project, the face classifier is enlarged by 10% of its minimum size. Increasing the *scale-factor* reduces the number of times the image is scanned, which increases the processing speed of the algorithm. However, a larger *scale-factor* reduces the accuracy of the algorithm, as the algorithm scans for less variation in facial size.

The Haar-object detection algorithm determines a potential ROI for an object by averaging all overlapping neighbouring candidate regions. The number of overlapping neighbouring candidate regions is averaged to form a ROI proportional to the confidence level of that ROI to contain the object of interest. The input argument *min_neighbours* specifies a minimum number of overlapping neighbouring candidate regions a fROI must be averaged from for it to be considered; otherwise, the fROI is rejected. In this project it was observed that most of the correctly detected fROI had more than 10 neighbouring candidate regions. From trial and error, the input argument *min_neighbours* was conservatively set to 3 neighbouring regions as it allowed detection of fROI with relatively low confidence level while minimizing the number of false positive face detections.

With the coordinates and size of each fROI, the Haar-face detection function also returns a number of neighbouring candidate regions that were averaged to form the corresponding fROI. Hence, if the Haar-face detection function returns multiple fROI for an image, the fROI with the largest number of neighbouring regions (in other words, the fROI with the highest confidence level) is selected. Table 5-1 summarizes the values of relevant corresponding input parameters of Haar-face detection function used in this project.

Table 5-1. Summary of input arguments of cvHaarDetectObjects function and their values used in this project

Input argument	Value
Cascade	"haarcascade_frontalface_alt.xml"
Flags	0
min_size	80 x 80
scale_factor	1.1
min_neighbours	3

5.3.1 Unstable face region of interest

Although the Haar-face detection algorithm correctly encapsulates a face within the detected fROI in the majority of frames, the fROI between consecutive frames in a given sequence of video appears too noisy and fluctuates even when the subjects are relatively stationary. Figure

5-3 shows an image formed by adding resulting frames from background subtraction of 18 consecutive frames selected from approximately 12 s of video data and then plotting the detected fROI of each respective frame on the final resulting image. Addition of resulting frames of background subtraction formed a white silhouette in regions where movement occurred due to difference in pixels intensity between frames, where as, the regions that were stationary between frames got cancelled and appeared black.

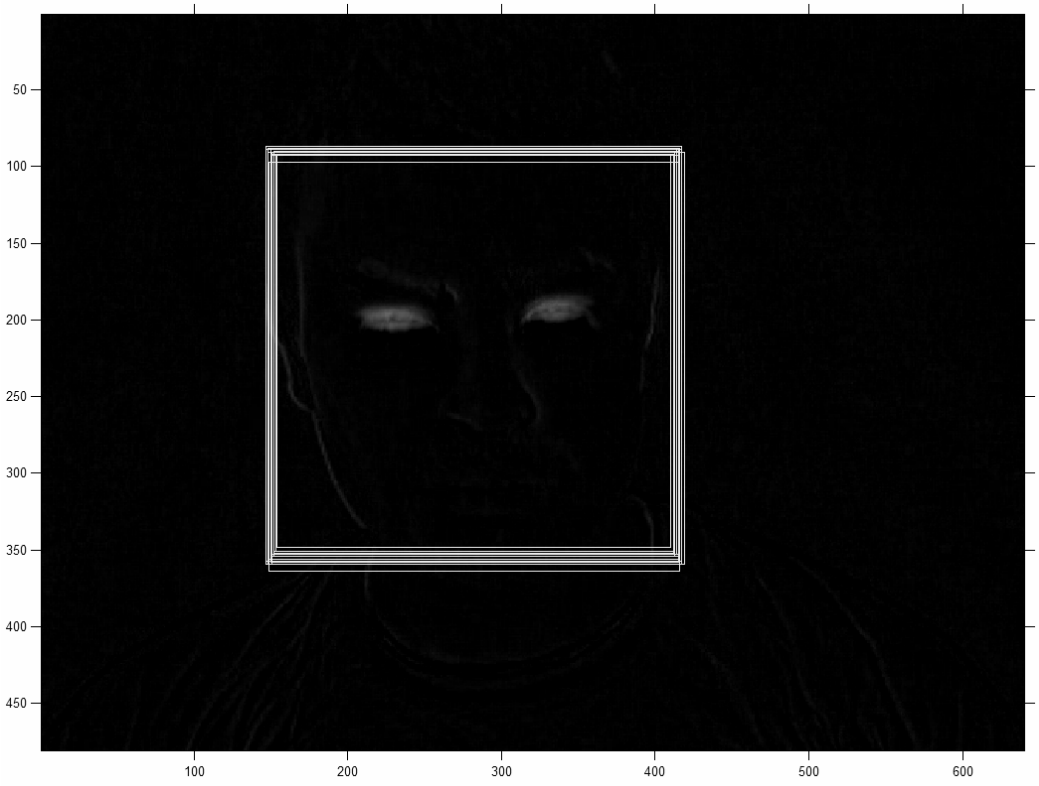


Figure 5-3. An image demonstrating the variation in detected fROI even when the subject remains relatively stationary. The image was formed by adding resulting frames of background subtraction of 18 consecutive frames selected over 12 s of video.

In Figure 5-3, there is a thin white silhouette outlining the edge of subject's face represents a small head movement and two large silhouettes at the eye region represents a large localized eyelid movements. Hence, over this 12 s of video, the subject made little head movements with multiple eye blinks. Comparing the thin white silhouette from small head movement relatively to the large distribution of fROI in the resulting images indicate that the localization and size of the fROI in consecutive frames varies substantially even when subjects remain relatively stationary.

Figure 5-4 shows plots of size, x_{min}, and y_{min} of fROI detected in each frame in 12 s of video used above, during which the subject's head appeared fairly still. From these plots, it was further evident that both size and position of fROI fluctuated, even when the subject remained relatively stationary.

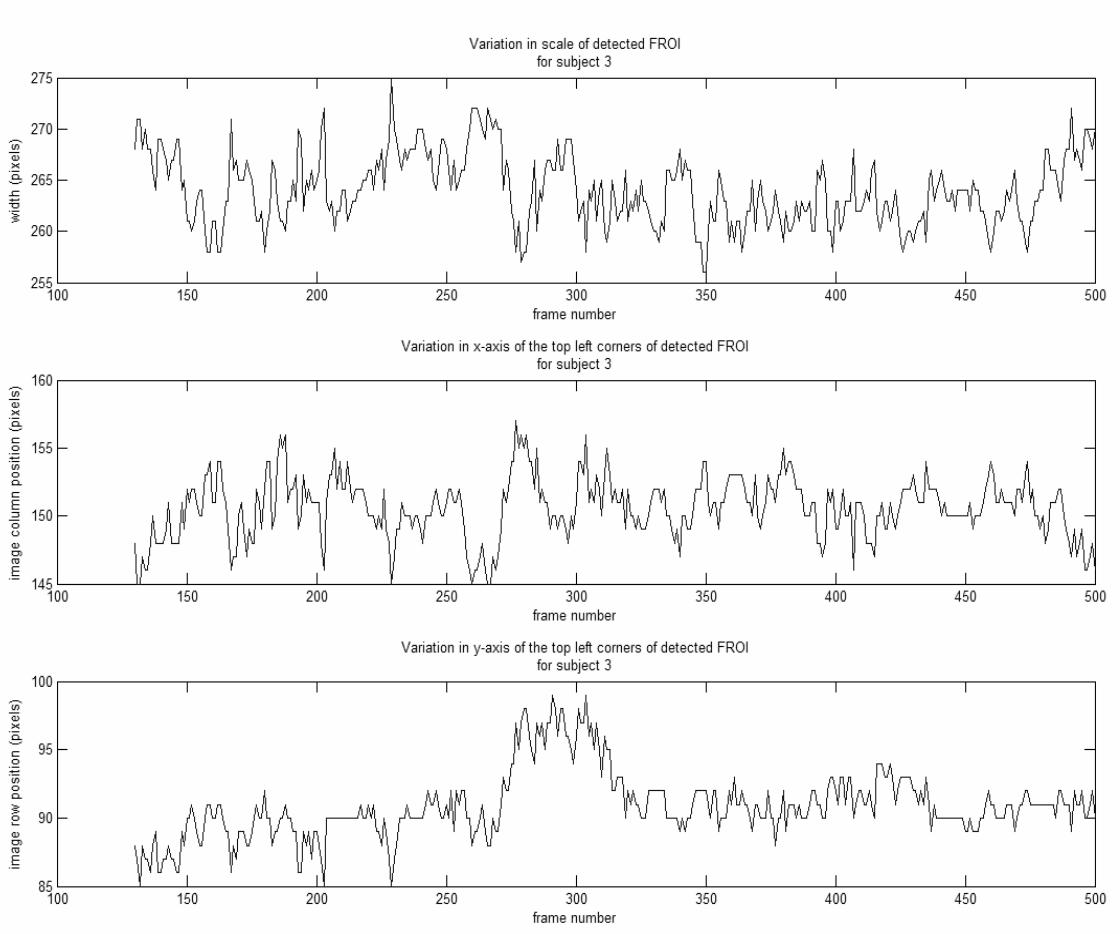


Figure 5-4. The size and position of detected fROI significantly fluctuated substantially in consecutive frames even when subject appeared relatively stationary.

Furthermore, sudden spikes in plot of size of fROI, as shown in frame number 928 in Figure 5-5, were also observed in some videos. These spikes were formed due to incorrect estimation of fROI in a frame rather than sudden head movements.

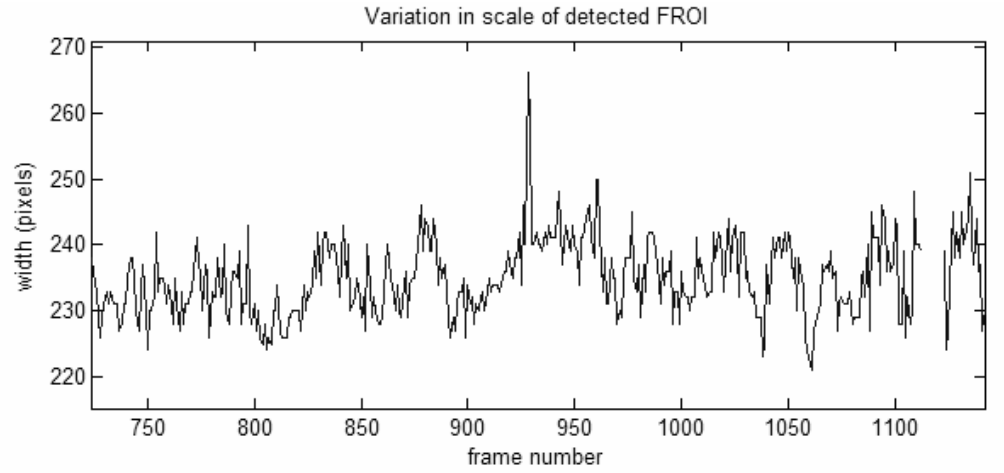


Figure 5-5. A plot of size of fROI for a sequence of frames in which a spike occurred at frame number 928 due to incorrect estimation of fROI by Haar-face detection algorithm.

In addition, the Haar face detection algorithm occasionally failed to detect the fROI as shown in between frame numbers 1112 to 1123 in Figure 5-5. In particular, the algorithm failed to detect the fROI in sequence of frames in which there were large movements due to head nod. Figure 5-6 shows nine spikes in y_{min} of fROI representing nine head nods performed by a subject during video data collection. Note that in the frames at the climax of each head nod, the Haar-face detection algorithm failed to detect the fROI.

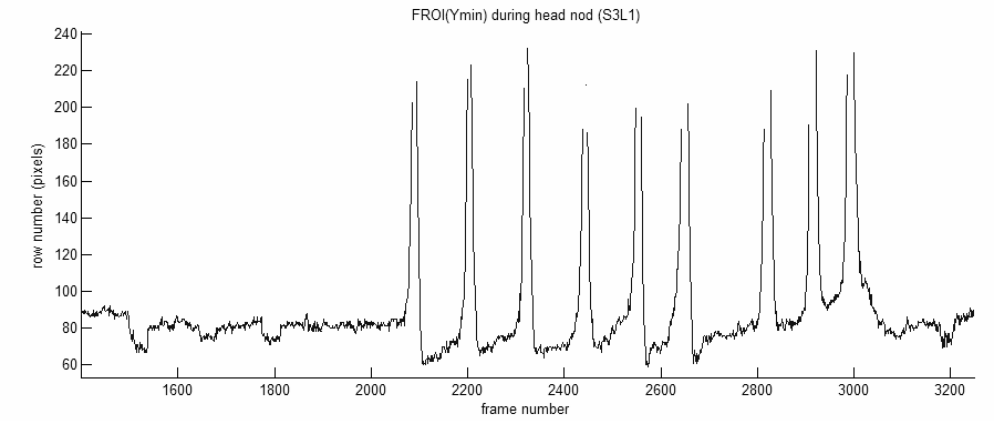


Figure 5-6. Plot of y_{min} of fROI. The 9 spikes represent the 9 head nods performed by the subject during video data collection. Note that the Haar-face detection algorithm failed to detect the fROI once the face becomes mostly invisible during downward movement during head nod and starts to detect the face again once the upright head posture is recovered.

Errors in the fROI detection affects the performance of the eROI localization (Figure 5-7) and eye closure measurement methods as both of these methods use parameters relative to fROI. In

this project, the eROI is derived with respect to the fROI. In addition, the centre of the fROI is used as a local stationary reference point for measuring eye closure. The details of these methods are discussed in later sections of this chapter. Due to the dependency of subsequent methods, the reduction of noise or jitter in fROI detected by the Haar-face detection method in consecutive frames of video was considered important to improve the performance of overall system.

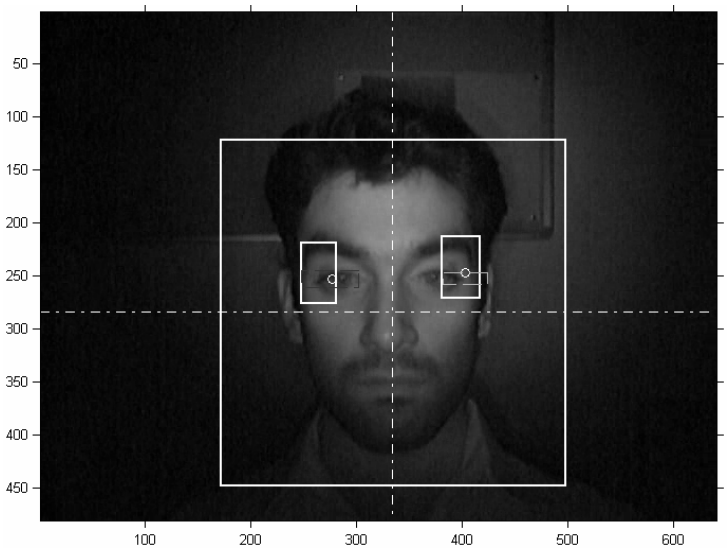


Figure 5-7. Example of incorrect estimation of eROI due to large size of the detected fROI.

5.3.2 Stabilizing the face region of interest

The estimated fROI in video must be filtered and tracked to reduce the jitters, spikes and failed detection. To achieve a stable fROI, the selected filter must be a low-pass filter and must be applied to each parameter in fROI to remove high frequency noise. Also tracking the parameters of the detected fROI in consecutive frames will reduce loss of fROI detection. The filter must also be causal[5] because the final lapse detection system must operate in real-time.

An exponentially-weighted moving-average (MA_{exp}) filter and a Kalman filter were both evaluated for filtering and tracking the parameters of the fROI. Both of these filters are causal and can perform low-pass filtering. These filters operate by taking the weighted value of past samples into account when estimating the present value of an input parameter. However, the Kalman filter is a more advanced filter than MA_{exp} as its adaptive filtering property dynamically adjusts its weighting on the values of past samples depending on the pattern of

[5] In causal filters the calculation of current state is fully based on past samples. Hence, these filters are applicable in real-time operations in which the post-processing is undesirable due to introduction of delays.

measured input values, whereas the MA_{exp} uses a constant weighting on past values. Hence, Kalman filter was used for stabilizing fROI detected by Haar-face detection method.

5.3.2.1 Kalman filter

The Kalman filter is a set of mathematical equations that provides an efficient recursive means to estimate the state of a process in a way that minimizes the mean squared error. This filter recursively conditions a current estimate based on past measurements and the process is repeated with the previous posterior estimates used for projecting the new *a priori* estimates. This recursive nature is one of the appealing features of the Kalman filter since it makes practical implementation more feasible (Zhu & Ji, 2005). A detailed description of Kalman filter is outside the scope of this thesis and only a very brief description is outlined in this section. Interested readers are directed to treatments such as Brown & Hwang (1997) and Welch & Gray (2006) for a more in-depth description of Kalman filter.

The general approach of a Kalman filter is to estimate the state $x \in \Re^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation

$$\mathbf{x_k} = \mathbf{A}\mathbf{x_{k-1}} + \mathbf{B}u_{k-1} + \mathbf{w_{k-1}}$$

Equation 5-1

The *n* x *n* matrix **A** in the difference-equation in Equation 5-1 relates the state of the previous time step to that of the current time step and is assumed to be constant. The *n* x *1* matrix **B** relates the optional control input u (scalar) to **x**.

The estimation is based on a measurement $\mathbf{z} \in \Re^m$, which is related to the process by

$$\mathbf{z_k} = \mathbf{H}\mathbf{x_k} + \mathbf{v_k}$$

Equation 5-2

In Equation 5-2, the *m* x *n* matrix **H** in the measurement equation relates state to measurement and is assumed to be constant. The random variables $\mathbf{w_k}$ and $\mathbf{v_k}$ in Equations 5-1 and 5-2 respectively represent process and measurement error. They are assumed to be independent of each other and zero-mean Gaussian distributed, thus $P(w) \sim N(0,Q)$ *and* $P(v) \sim N(0,R)$, where *Q* and *R* are the process noise covariance and measurement noise covariance, respectively, and are assumed to be constant.

The Kalman filter algorithm can be divided into two processes: the time update process and the measurement update process (Welch & Gray, 2006). The Kalman filter uses these two processes in the form of feedback control system. The operation of the Kalman filter algorithm with the equations of both time update and measurement update processes is outlined in Figure 5-8. The equations in the time update process are used for projecting forward (in time) the current state $\hat{x}_k^-$ and error covariance estimates P_k^- to obtain the *a priori* estimates for the next time step. The equations in the measurement update process are responsible for the feedback part of the Kalman filter algorithm, where the latest measurements value, $\mathbf{z_k}$, are incorporated into *a priori* estimate to obtain the new estimated state $\hat{x}_k$ and an improved *a posteriori* estimate P_k. As the actual state is unknown, the error covariance can only be estimated. $\mathbf{K_k}$ is the Kalman gain matrix, which controls the reliance that the state estimate places on the latest measurement as against the past knowledge.

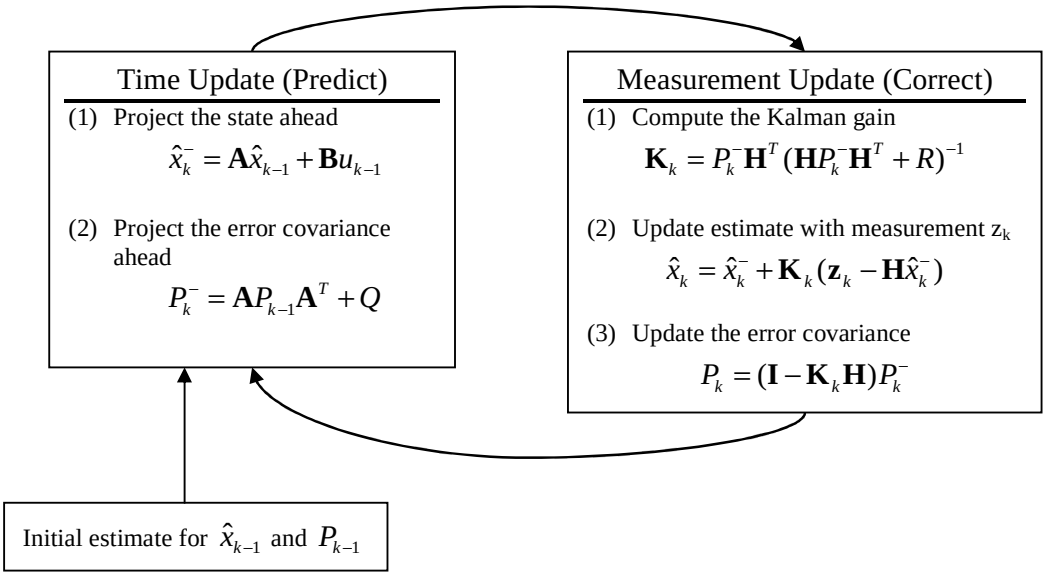


Figure 5-8. The Kalman filter algorithm with equations of time update and measurement update (Welch & Gray, 2006).

5.3.2.2 Filtering the Haar face variables

The top-left coordinate pair (x_{min}, y_{min}) and a single size variable of fROI are likely to be highly correlated. To filter these variables a very simple model is adopted for the Kalman filter. Here, for convenience, only the x_{min} coordinate of Haar face variable is considered; the same process is applied to filter the rest of the variables. The state vector used is

$$\mathbf{x} = \begin{pmatrix} x_{\min} \\ \dot{x}_{\min} \end{pmatrix}$$

No control input is incorporated, so Equation 5-1 for the time step process becomes:

$$\mathbf{x_k} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \mathbf{x_{k-1}} + \mathbf{w_{k-1}}.$$

Without loss of generality, the Δt was set to same as the sampling interval of 1 ($\Delta t = 1$) so that the $\dot{x}_{\min}$ state variable is scaled according to the sampling rate. Only the state variable $x_{\min}$ is actually observed with the implemented Haar face software, which results in Equation 5-2 becoming

$$\mathbf{z_k} = \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{x_k} + \mathbf{v_k}.$$

It remains to choose values for *Q* and *R* and to initialize the filter. *Q* represents the covariance of the process (the model); *Q* affects the smoothness of the estimate and has been set by trial-and-error. *R* represents the covariance of the measurements and is a scalar in this case. Choosing a large value for *R* implies a large uncertainty in the measurements and a less responsive filter output. The quantity $(\mathbf{H}P_k^{-}\mathbf{H}^T + R)$ is scalar and so the inverse during the update of estimate with the measurement value $\mathbf{z_k}$ in the measurement update process in Figure 5-8 becomes a simple division. Finally, given the simple model adopted and lack of genuine prior information, a simple approach has been used to initialize the Kalman filter with

$$\hat{\mathbf{x}}_1 = \begin{pmatrix} x_{\min,1} \\ 0 \end{pmatrix}$$

$$P_1 = \begin{pmatrix} \in & \in \\ \in & \in \end{pmatrix}$$

where $\in$ is a very small value (e.g., 10^{-4}).

5.3.3 Evaluation of filtered face region of interest

The Kalman filter considerably reduced the high frequency noise in fROI. As a result, there was less variation in position and size of fROI between consecutive frames in which the subject remained relatively still. For example, Figure 5-9 shows the low-pass filtering effect of the Kalman filter on size x_{min} and y_{min} of fROI detected in the same sequence of frames as

shown in Figure 5-4. The Kalman filter output (solid black line) of each parameter is much smoother than the raw data. Visual inspection also confirmed that the fROI in the video sequence is much more stable.

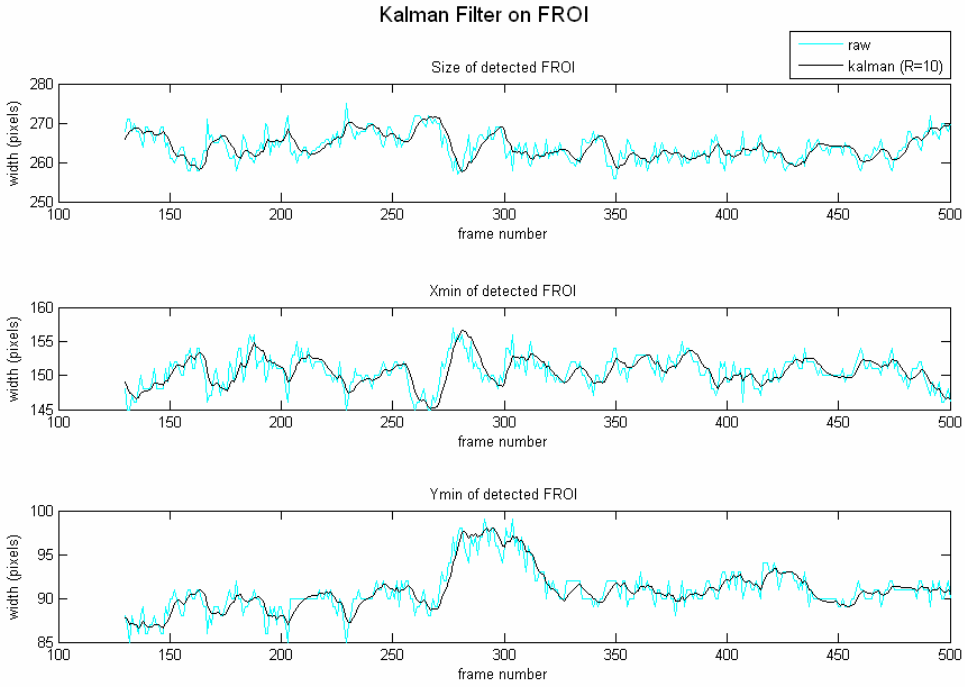


Figure 5-9. Result showing the removal of high frequency noise in x_{min}, y_{min}, and size parameters of fROI by applying Kalman filter with the measurement error covariance set to 10.

The lowpass filtering nature of the Kalman filter also filtered out any large error in estimation of fROI. For example, the incorrect estimation of fROI's size parameter in frame 928, which was identified by a spike in Figure 5-5, was corrected by the Kalman filter as shown in Figure 5-10. More importantly, the tracking or predictive nature of the Kalman filter provided an estimation of fROI in frames when the Haar-face detection method failed to detect one. For example, in frames 1112 to 1123 in Figure 5-10, the Haar-face detection method failed to detect fROI but the Kalman filter predicted the fROI in those frames, albeit poorly, based on past temporal information. Although the confidence level in accuracy of fROI detection is relatively low in these frames, the tracking ability of the Kalman filter is particularly useful for identifying the fROI under head movement such as head nod. As shown in Figure 5-11, where, during all nine head nods performed by the subject, the Kalman filter correctly predicted the y_{min} of fROI even when the Haar-face detection algorithm failed to detect fROI.

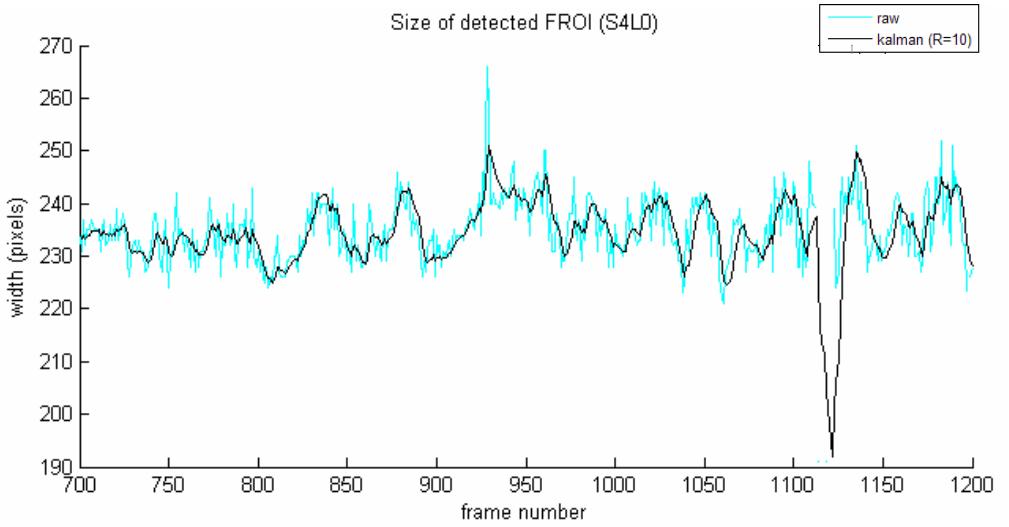


Figure 5-10. The spike at 928 is smoothed and also the loss of fROI detection between frames 1112 to 1123 is filled due to tracking.

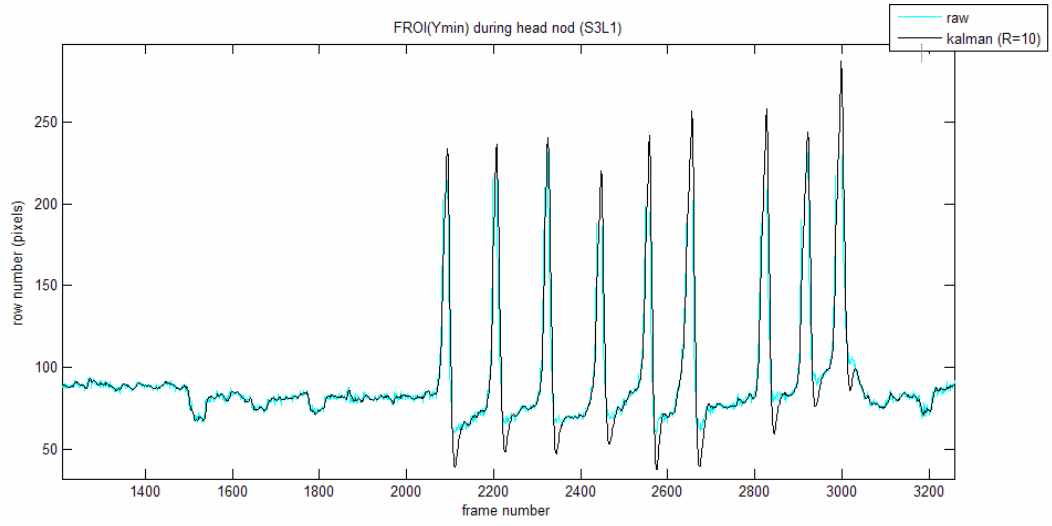


Figure 5-11. Plot of y_{min} coordinate showing the tracking of fROI by Kalman filter during head nod.

The downside of Kalman filtering is the time lag in its output response, as with all causal filters. In Figure 5-12, the lag in response of Kalman filter is compared with the response of a lag-less non-causal central-difference-based moving-average filter with sampling window of 10 samples. In this example, comparing the corresponding peaks in responses of the two filters showed a worst peak-to-peak delay of 5 frames. When acquiring an image at 30 fps, a delay of 5 frames equates to 167 ms. This lag will usually have minimal effect on lapse detection, which usually appears as a series of slow eyelids and head movements over a much longer period of

time. Hence, each parameter of fROI is filtered through the Kalman filter before being used by subsequent methods in the current system.

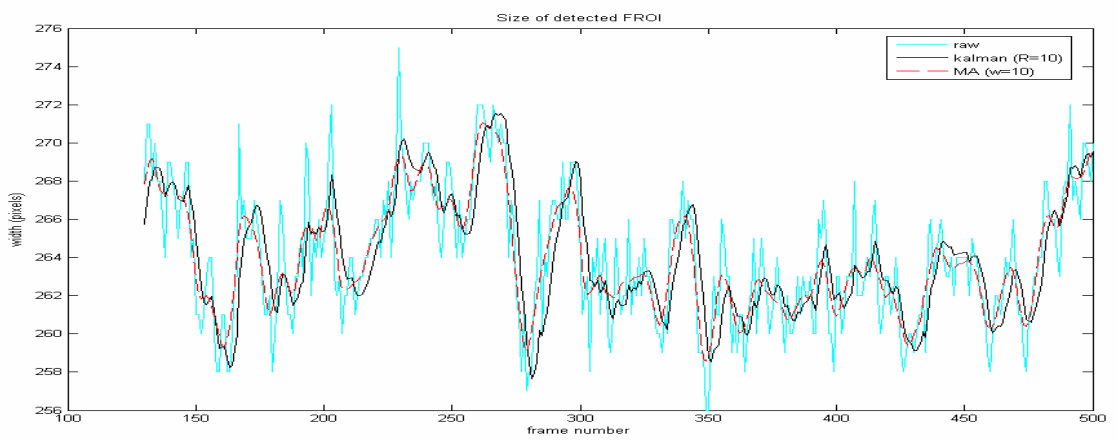


Figure 5-12. Plot comparing the lag in Kalman filter response (dark solid line) against the non-causal central-difference-based moving-average filter (dashed line) without lag in its response.

5.4 Anthropomorphic eye region of interest localization

Once the fROI within an image is determined, the next step towards measuring eye closure is to localize eye regions of interest (eROI). An eROI defines a region within an image in which an eye will be present. Left and right eROIs are outlined by separate rectangles. However, unless the side of an eROI is specified, the term "eROI" represents both left and right eROIs. A common procedure is applied to both eROIs in most cases. An eROI is derived by scaling the parameters of detected fROI with predefined proportional anthropomorphic constants (as described in section 2.3.3.2). The constants are derived by analysing the annotated position of eyes relative to the fROI detected in the annotated image database.

5.4.1 Derivation of proportional constants between eROI and fROI

The proportional constants for the rectangular eROI parameters relative to fROI were derived by forming a rectangular ROI which encloses the distribution of centre of the annotated visible eye (COE_a) relative to fROI in annotated frames. Figure 5-13, shows an example of left and right COE_a (marked by 'x') in one of the annotated frames.

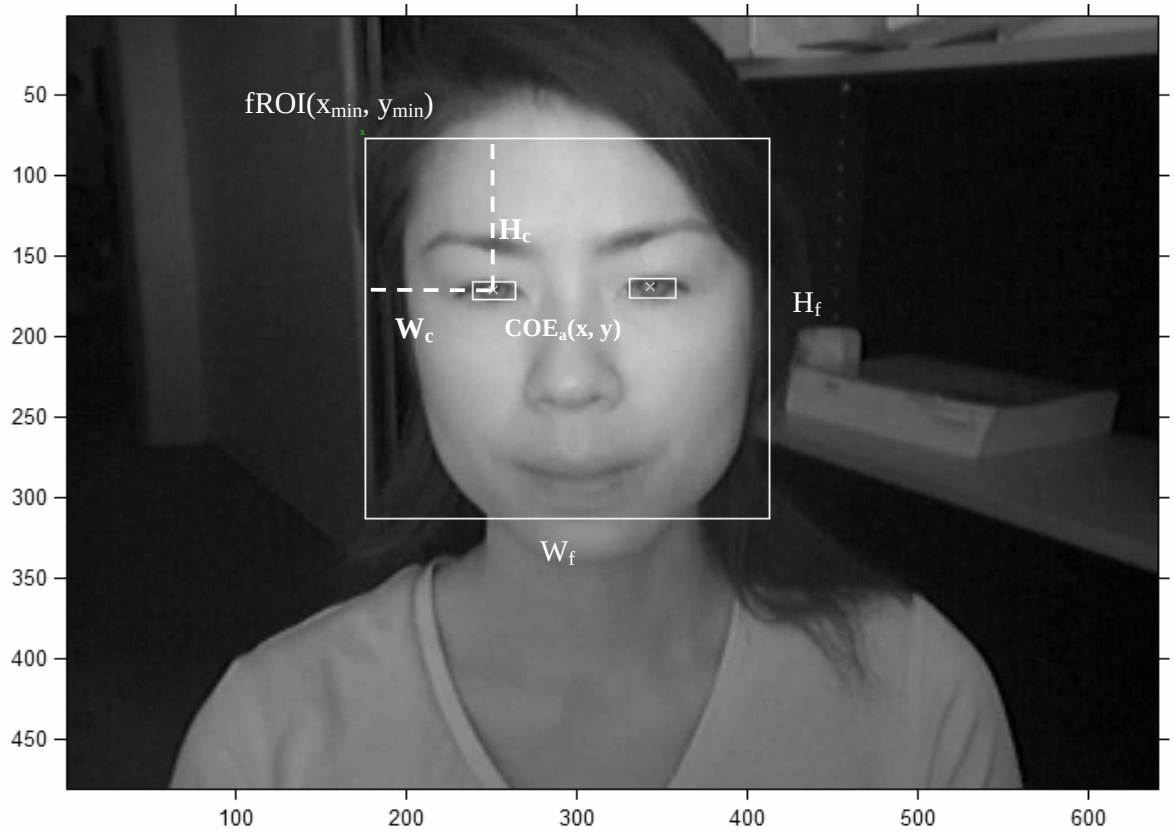


Figure 5-13 An image from the annotated image database showing the referencing of right COE_a relative to top-left corner (fROI(x_{min}, y_{min})) and proportional to the size *(W_f, H_f)* of detected fROI.

The distribution of COE_a relative to fROI was formed by plotting the distance between the COE_a and the top-left corner position, fROI(x_{min}, y_{min}), of fROI as proportional to the size of the fROI in each annotated images. Equations 5-3 and 5-4 show the calculation of the proportional distance in x-axis *(W_c)* and y-axis *(H_c)*.

$$W_c = \frac{COE_a(x) - fROI(x_{min})}{W_f}$$

Equation 5-3

$$H_c = \frac{COE_a(y) - fROI(y_{min})}{H_f}$$

Equation 5-4

Note: Since fROI is a square in this project, W_f and H_f are equal.

The W_c and H_c for both left and right eyes from each reference images in the annotated database were plotted to form a COE_a distribution relative to fROI, as shown in Figure 5-14[6]. The two clusters in Figure 5-14 represent the population distribution of respective left and right COE proportional to fROI. Hence, it can be assumed that for any given fROI, the left and right COEs will most likely be located within the area bounding the respective clusters, represented by the two respective rectangles that enclose the left and right clusters in Figure 5-14. These proportional eROI were denoted as $eROI_f$. The width and height of an $eROI_f$ are defined by the distance between extremes of W_c and H_c, respectively. The sides of the $eROI_f$ were extended by 10% to increase the tolerance for COE localization. Since the fROI were detected by Haar-face detection algorithm and filtered by Kalman filter prior to deriving the $eROI_f$, any noise in fROI were accounted within the COE_a distribution.

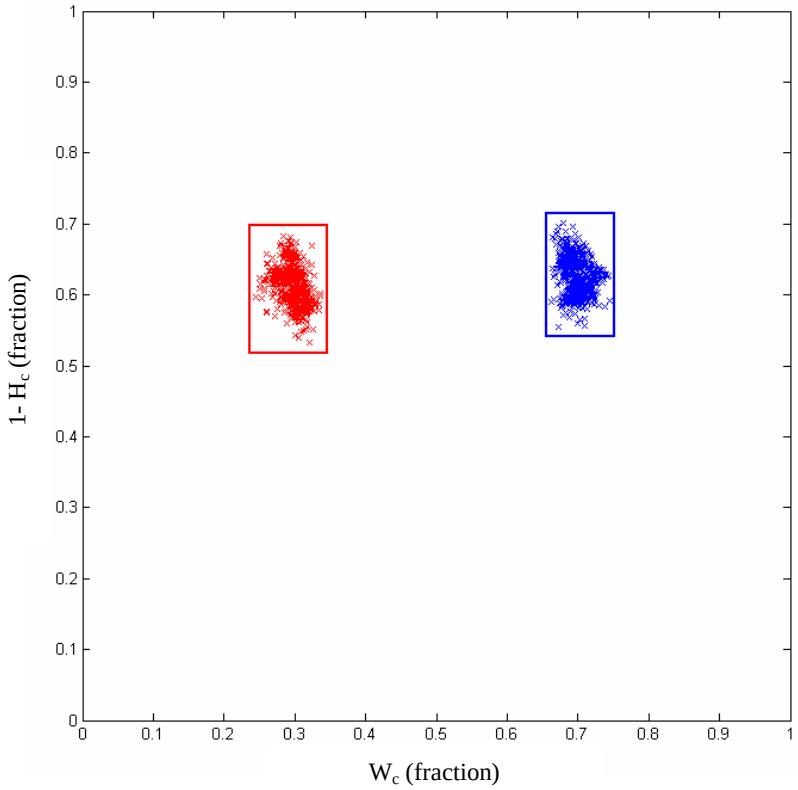


Figure 5-14. The left and right COE_a of images in the annotated image database plotted as proportional to corresponding fROI detected in the images. The two clusters of proportional COE_a represent the population distribution of COE relative to fROI. The rectangles ($eROI_f$) that enclose each cluster represent the respective left and right eROI that is proportional to fROI.

[6] Note that the H_c in Figure 5-14 are inverted. The rows of the image are indexed from the top.

Each $eROI_f$ is defined by its top-left corner position $eROI_f(x_{min}, y_{min})$, width $eROI_f(w)$, and height $eROI_f(h)$, extracted directly from Figure 5-14. Table 5-2 lists the value of parameters of both left and right $eROI_f$ proportional to fROI. Note that unlike inverted H_c values in Figure 5-14[6], Table 5-2 presents the actual values.

Table 5-2. Values for the left and right $eROI_f$ proportional to fROI as extracted from Figure 5-14.

	$eROI_f(x_{min})$	$eROI_f(y_{min})$	$eROI_f(w)$	$eROI_f(h)$
Left eye	0.24	0.32	0.09	0.15
Right eye	0.66	0.29	0.08	0.15

For any given fROI, an eROI was derived by using the proportional $eROI_f$. For example, for a fROI whose top-left corner position, width, and height parameters were defined by $fROI(x_{min}, y_{min}, w, h)$, the corresponding eROI were derived as shown in Equations 5-5, 5-6, and 5-7.

$$eROI(x_{min}) = (fROI(x_{min}) + (eROI_f(x_{min})) \times fROI(w))$$

Equation 5-5

$$eROI(y_{min}) = (fROI(y_{min}) + (eROI_f(y_{min})) \times fROI(h))$$

Equation 5-6

$$eROI(w) = eROI_f(w) \times fROI(w) \quad eROI(h) = eROI_f(h) \times fROI(h)$$

Equation 5-7

Figure 5-15 shows an example of estimated left and right eROIs (dashed lines) for the annotated image in Figure 5-13. The COE_a is also marked in the image and falls inside the estimated eROI. Figure 5-16 shows eROI estimation on an image that was not part of the annotated image database and, therefore, was not used for deriving the $eROI_f$. Note that the eROIs do not fully encapsulate the corresponding eyes because the eROI is derived to determine the position of the COE (section 5.6.2) and only needs to enclose the region in which the COE is most likely to be present. Although extending the eROI will increase the likelihood of enclosing the COE, it will also increase the search area, hence, making the system computationally inefficient (section 5.6.1).

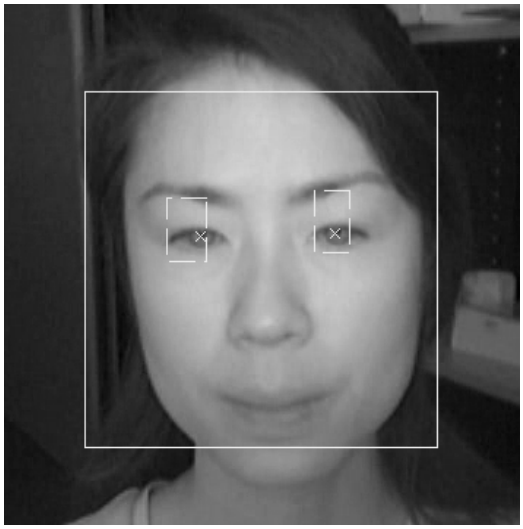


Figure 5-15. An example of estimated left and right eROIs (dashed lines) (with the corresponding annotated centre of visible eye marked with 'x' inside the eROIs) for the annotated image in Figure 5-13.

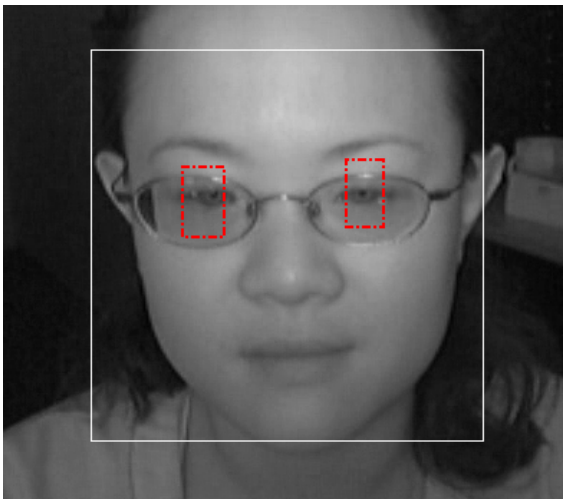


Figure 5-16. Example of non-annotated image displaying the eROI estimated based on anthropomorphic method.

5.4.2 Evaluation of eROI localization

Randomly selected frames from videos of each subject in the reference database were visually analysed to assess the inclusion of eye within the localized eROI. The eROI and fROI were marked on each analysed frame. The frames for visual inspection were randomly selected based on different head orientations of subjects, which included frontal face images with upright head posture and images in which their head was turned to up, down, and sideways gazes with various angle of head rotation.

In all annotated frontal face images in which the fROI was accurately detected, the COE_as were included within the corresponding estimated eROIs. Further visual inspection of videos showed that eROI included the COE in most frontal face images even under fluctuating fROI output. Correct detection of eROI in frontal face images was particularly important because the subjects are expected to be facing straight ahead when they are drowsy or about to have microsleeps while performing visuomotor tasks. Furthermore, as shown in Figure 5-17, the estimated eROI included parts of an eye in some frames even when the subject's head was turned. It was observed that if the planar position of detected fROI changed relative to direction of head rotation, the eROI were correctly estimated. However, the Haar-face detection algorithm was only able to accurately detect the fROI for small head rotations.

Figure 5-17. Examples of correct eROI estimation in frames in which subject's head were rotated in various directions.

Figure 5-18 show two examples in which inaccuracy of fROI detection in frames with large head rotation results in incorrect estimation of eROIs. In Figure 5-18(a), the detected fROI completely misses the face, so obviously the eROIs were also incorrectly estimated. However, in Figure 5-18(b), although the subject's face is enclosed within the fROI, size of the fROI is relatively larger than fROI detected when the subject was facing straight ahead. The larger fROI resulted in incorrect estimation of eROI because the proportional anthropomorphic

constants that were used to estimate the eROI relative to fROI were derived based on fROI detected in frontal face images. As expected, it was observed that the reliability of the anthropomorphic proportional constants based eROI localization method is dependent on the precision of fROI detection.

(a) (b)

Figure 5-18. Examples where the inaccurate detection of fROI results in incorrect estimation of eROI. (a) The detected fROI completely misses the face. (b) The large size of fROI results in incorrect estimation of eROI as the eROI is directly proportional to the size of fROI.

In conclusion, the eROI was able to reliably enclose central parts of an eye when a subject is facing straight ahead naturally without any external head constraints. However, as the precision of fROI deteriorates with increase in angle of head rotation, the localization of eROI becomes less reliable. Research is being carried out in applying angular Haar-object classifiers to detect rotated objects (Barczak, 2005), which could be incorporated in this project in future to make the fROI detection more robust under larger angular head orientation. However, correct estimation of eROI in frontal face images was acceptable at this initial stage of this project because subjects are most likely to be facing straight ahead during the event of drowsiness and microsleep.

5.5 Performance evaluation method

Unlike the face and eye region of interest detection methods, whose performance was determined visually, the performance of the eye-feature detection methods presented in the next few sections were quantitatively evaluated. This section describes the error metrics used for determining the performance of the eye feature detection methods.

The error of an eye-feature detection method was calculated by taking the difference between the estimated and annotated (section 4.2) image coordinates of the eye-feature. The statistical parameters of the estimation error and its magnitude were then analysed to determine the performance of the eye-feature detection method. For each eye-feature detection method, the mean and standard deviation (SD) of errors in the estimated eye-feature position in each annotated frames in the reference database were calculated. In addition, differences in the median error magnitude and the 90^{th} percentile error magnitude between subjects were also determined. The median and 90^{th} percentile of the error magnitude were used instead of mean value because the distributions of the error magnitudes for all of the eye-feature detection methods were found to be skewed towards a large number of eROI with low error magnitudes as oppose to a few eROIs with relatively high error magnitudes. Hence, the median error magnitude represents the general performance of an eye-feature detection method in the majority of eROIs of a subject, whereas the 90^{th} percentile error magnitude represents a measure of the worst-case performance. Furthermore, the means and SDs of the medians and the 90^{th} percentiles of error magnitude for each subject were calculated to derive the overall general and worst-case performances of each eye-feature detection method, respectively. The performance of eye-feature detection methods were further analysed for the group of subjects who wore glasses verses the group who did not, so as to evaluate the degradation in the method due to presence of glasses in front of the eyes.

5.6 Centre of eye detection

The vertical integral projection (*VIP*) function (section 5.7.2) offers a simple and computationally efficient method for detecting vertical positions of eyelids (Hua Gu et al., 2003; Zhou & Geng, 2004). However, from literature review and preliminary experiment, it was realised that the *VIP* function based eye feature detection methods performed reliably only when the eROI were small and tightly enclosed an eye to exclude other facial features. The performance of the method deteriorated for larger eROI because of the presence of features like the eyebrow and the eye-glasses. As shown in Figure 5-19, the eyebrow projects similar *VIP* characteristics as the eye, which may result in false detection of the eyebrow position as the eye position. Since it cannot be guaranteed that the estimated eROI will include only an eye and exclude any additional features surrounding the eye, it was important to distinguish an eye from other features within the eROI before estimating the positions of eyelids.

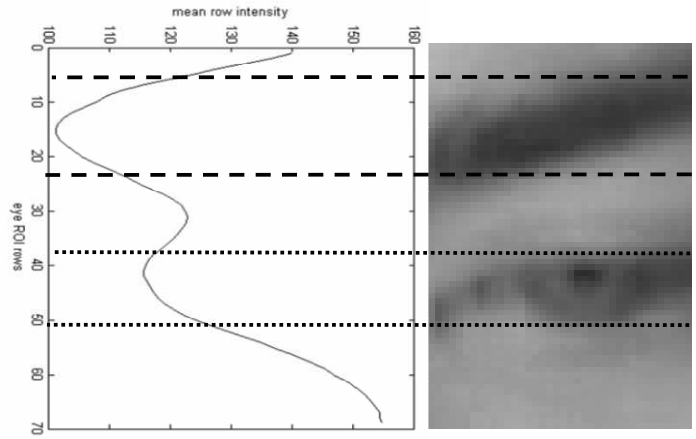


Figure 5-19. An illustration of the similarity in the pattern of image intensity of the eyebrow and the eye in vertical integral projection of an eROI.

To distinguish the position of the eye from non-eye features within an eROI, the position of a known eye feature must be detected. The centre of eye (COE) position was considered to be a good eye feature to distinguish an eye, as it can be detected by using a simple template matching method. The COE represents the stationary centre of the visible parts of the eyeball (section 4.2.2) and should not be confused with the pupil. In addition to distinguishing an eye from the other features, the estimation of vertical position of COE also helps in detecting the position of eyelids in *VIP* because the COE lies vertically between upper and lower eyelids and horizontally aligns with the apices of the eyelids.

Detecting the COE using the template matching method involves formation of the eye-template and then cross-correlation of the eye-template with the eROI within an image. Initially an eye-template that encodes the image properties of the eye was formed. This eye-template was cross-correlated with sub-images extracted based on each pixel within an eROI to form a cross-correlation matrix of same size as the eROI. The cross-correlation matrix is further refined based on *a-priori* knowledge about likely the position of the COE. Finally, the position with the highest correlation coefficient within the refined cross-correlation matrix was selected as the corresponding COE position within the eROI. This procedure is discussed in more detail in next few subsections.

5.6.1 Forming eye template

An eye template (T) is formed by encoding the distinct elliptic shape of the eye and the contrast in image intensity within the eye features and its surroundings. Figure 5-20 shows eye templates that were formed and tested. Two main groups of eye templates were formed. For each group, separate left and right eye templates were initially used for detecting the COE in respective sides, then the left and right templates were averaged to form an overall mean eye template. Initially a sub-image that tightly enclosed an open eye of Subject-6 was cropped with the annotated COE positioned at the centre of the sub-image to form a simple single eye template. The performance of the single eye template was used as a base-line measurement of COE detection.

The second group of eye templates were formed by taking the average intensity of sub-images of fully opened eyes extracted from a set of 378 selected annotated frames of all nine subjects in the reference database. The images in row two, three, and four in Figure 5-20 shows the three different sets of mean eye templates trialled in this project. The selected frames of mean eyes comprised three frames for each of the seven different gaze directions (section 4.2.1) under both daylight and dark lighting conditions. These templates are likely to correlate better with wide range of subjects in the general population because they encode variability between subjects.

The mean eye template (T_W), shown in the second row in Figure 5-20, was selected to encode both the eyebrow and eye within an eye template with COE centrally located. For $T_{Woffset}$, shown in the third row, the COE was offset. The large sizes of the wide mean eye templates T_W and $T_{Woffset}$ led to very slow computation (approximately 0.2 fps). To improve the computational speed, a smaller mean eye template (T_S), as shown in the fourth row in Figure 5-20, was formed.

Evaluation of the eye templates showed that applying an average eye template of left and right eye templates to both left and right eROI generally performed better than applying separate left and right eye templates to the corresponding eROI. Table 5-3 shows the error in COE detection while using corresponding averaged eye templates shown in the fourth column in Figure 5-20. Although $T_{Woffset}$ had the lowest error magnitude in COE estimation, the mean eye template T_S in the forth row and forth column in Figure 5-20 was selected as the eye template to be used in this project because its error was only fractionally higher than $T_{Woffset}$ eye template while its

computational speed for template matching procedure was substantially better at 1 fps (i.e., 5 times faster than the two wider mean eye templates).

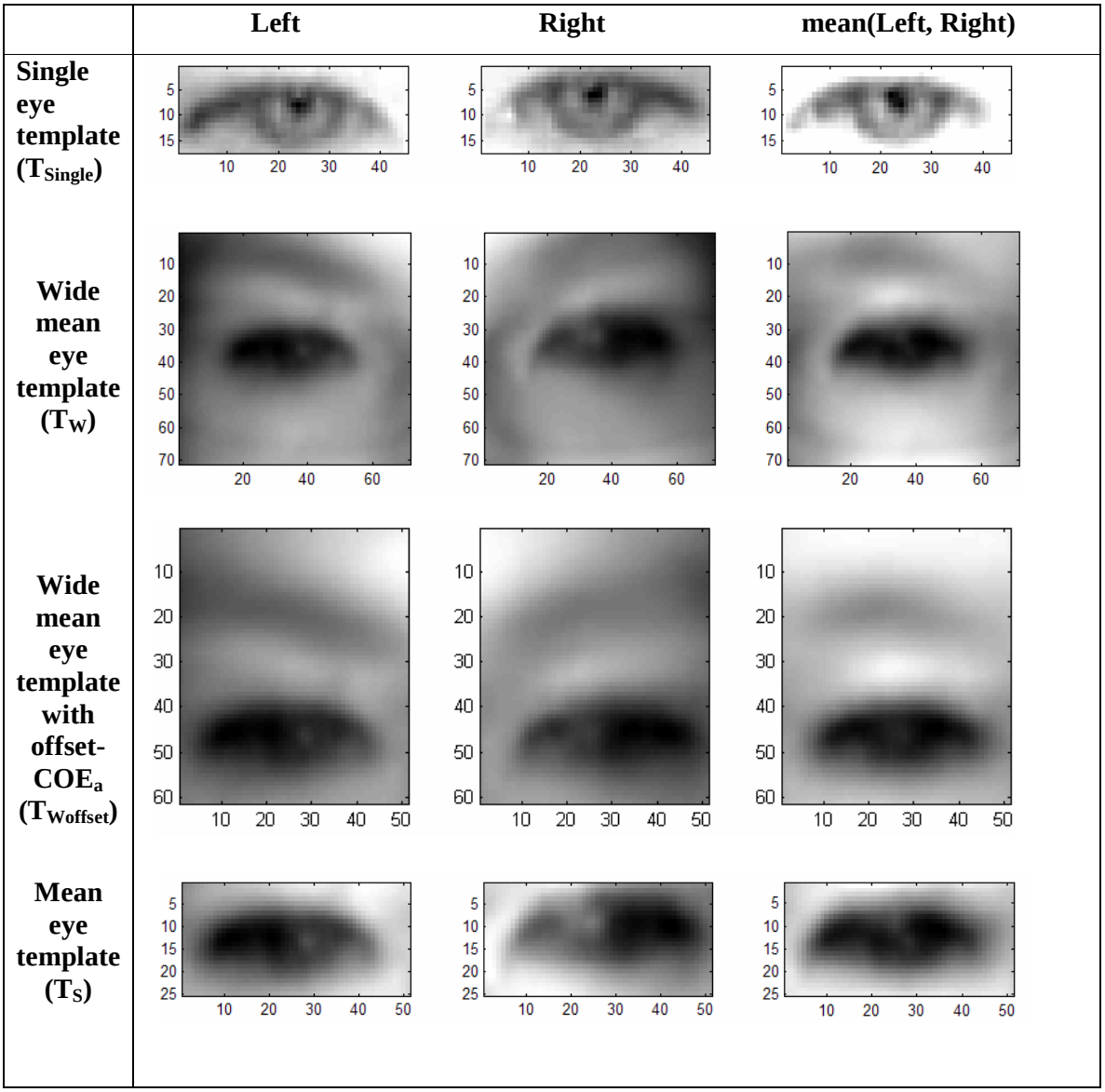


Figure 5-20. The eye templates formed and analysed for COE detection through template matching method.

Table 5-3. The mean 90th percentile error of the nine subjects in detection of COE by various eye templates.

Eye templates	mean 90th percentile error in COE(y) (pixels)
$\mathbf{T_{Single}}$ **(base-line)**	19.90
$\mathbf{T_W}$	9.71
$\mathbf{T_{Woffset}}$	5.65
$\mathbf{T_S}$	6.04

The T_S eye template was further altered to improve its performance. Performance analysis of frames showed that in some of the frames with horizontal and vertical gazes, the centre of dark iris was being falsely detected as the COE. This false detection was due to better correlation of the eye template with the dark intensity of iris then with the lighter intensity region at the COE due to presence of sclera when a subject is looking away from the camera. In addition, the NIR illumination source that was placed in front of the subject during video recording was being reflected of the centre of the spherical eyeball making the COE region even lighter in intensity, as shown in Figure 5-21.

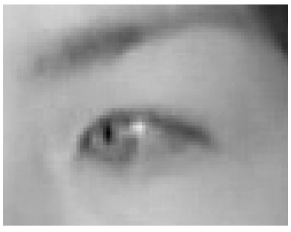


Figure 5-21. An image showing the reflection of the NIR illumination source off the sclera during the sideways gaze direction.

To make the COE detection with the mean eye template independent of the iris position, an elliptic "don't care" region that masks the visible eye was defined at the centre of eye template, as shown in Figure 5-22. During the calculation of correlation coefficient between the eye template and sub-images, the pixels under the "don't care" mask are ignored. The "don't care" region for the overall mean eye template was defined by an ellipse with radius (minor, major) = (3, 12) and centre (x, y) = (26, 13). The mean 90th percentile COE(y) estimation error for the T_S eye template improved from 6.04 pixels to 4.86 pixels when the "don't care" mask was added.

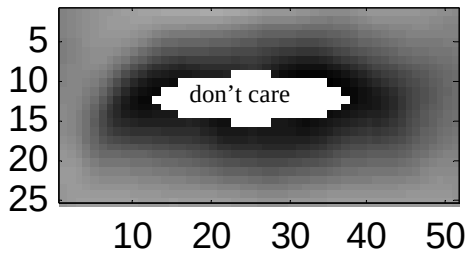


Figure 5-22. The mean eye template with don't care elliptic region set at the centre to improve the COE detection in the frames with sideways gaze direction.

5.6.2 Correlation matrix

Each pixel within the eROI has the potential be the COE. Hence, for each pixel positioned at (*x*, *y*) within the eROI, a 2D correlation coefficient was calculated between an eye template (**T**) and a sub-image ($\mathbf{S_{xy}}$) of same (N x M) size as **T** extracted from the image. Each $\mathbf{S_{xy}}$ is extracted so that the pixel at (*x*, *y*) position within the eROI corresponds to the annotated COE_a in **T**. The correlation coefficient (r_{xy}) for a pixel within the eROI is defined as

$$r_{xy} = \frac{\sum_n \sum_m \left(\mathbf{S_{xy}}(n,m) - \bar{S}\right)\left(\mathbf{T_{xy}}(n,m) - \bar{T}\right)}{\sqrt{\left(\sum_n \sum_m \left(\mathbf{S_{xy}}(n,m) - \bar{S}\right)^2\right)\left(\sum_n \sum_m \left(\mathbf{T_{xy}}(n,m) - \bar{T}\right)^2\right)}}$$

Equation 5-8

where $\bar{S}$ and $\bar{T}$ are the overall means of $\mathbf{S_{xy}}$ and **T** respectively. The correlation coefficient is a good metric for comparing two images because it provides a normalized comparison between two images and eliminates any bias due to the changes in intensity levels within input image. Also, the r_{xy} provides a convenient quantitative index which approaches 1 when a sub-image matches with the eye template image. The calculation of r_{xy} for each pixel positioned at (*x*, *y*) within the eROI forms a correlation matrix (**CC**) of same (X x Y) size as the corresponding eROI as illustrated in Figure 5-23.

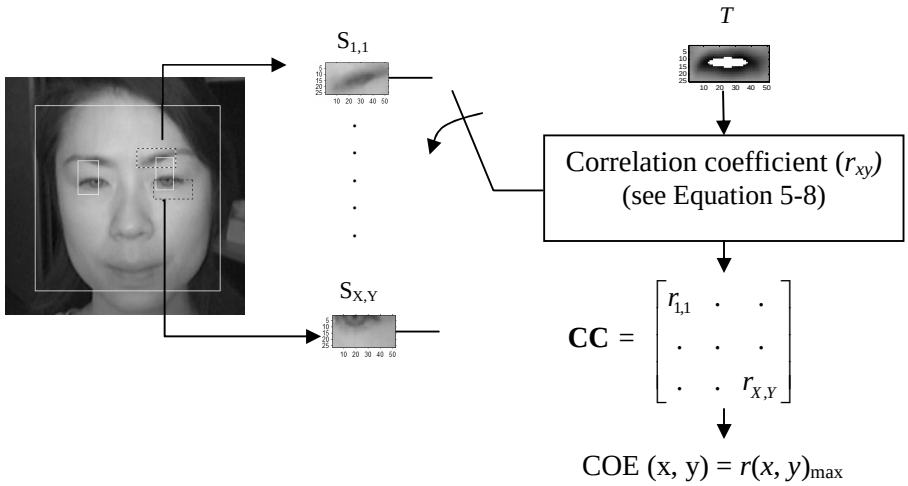
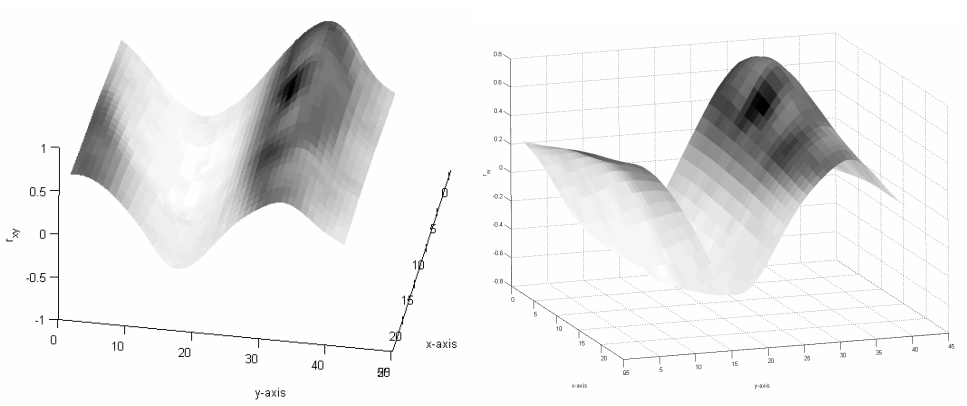


Figure 5-23. Steps involved in template matching method to form a correlation matrix (CC) for an eROI in which the position of with the highest correlation coefficient (r_{xy}) is selected as COE position within the eROI.

Figure 5-24 shows 3D mesh plot (from two different viewing angles) of **CC** formed for the right eROI of the subject shown in Figure 5-23. The higher *r* value in the **CC** shown in Figure 5-24(a) suggests that the eye template correlates better with the darker intensity such as eyebrow and eye region. In contrast, the eye template does not correlate well with the lighter skin region as suggested by the lower *r* value in that part of **CC**. In Figure 5-24(b), it is apparent that *r* in the **CC** peaks close to true COE and is much higher than the dark eyebrow region. The higher *r* towards the true COE is due to better correlation of the elliptic shape of an eye with the eye template.



(a) (b)

Figure 5-24. 3D mesh plot of the correlation matrix formed in Figure 5-23 with the superimposed EROI sub-image.

5.6.3 Initial performance of COE detection

The performance of COE detection method was determined by first analysing the distribution of estimated COE error in all subjects and then analysing the error magnitude in each individual subject. The difference between x and y coordinates of the estimated and the annotated COE was calculated and the absolute of the difference was taken. The resolution of an average eye in the reference database was calculated to be 20 x 45 pixels. Hence, in worst-case scenario, error in COE detection should be less than ±10 pixels in y-coordinate and ±22 pixels in x coordinate to reliably distinguish the eye from non-eye features in an eROI. The initial performance analysis of a developed COE detection method showed encouraging results and also helped to identify limitations in the method, which were further corrected.

Figure 5-25 shows the error distribution of x and y coordinates of estimated COE. The mean and SD of the error distribution of the x and y coordinates was 1.1 ± 4.6 pixels and 1.2 ± 6.7 pixels, respectively. In Figure 5-25, the error in estimation of the COE appears to have normal distribution with low mean error of close to a single pixel in both x and y coordinates. The low mean error was an encouraging result.

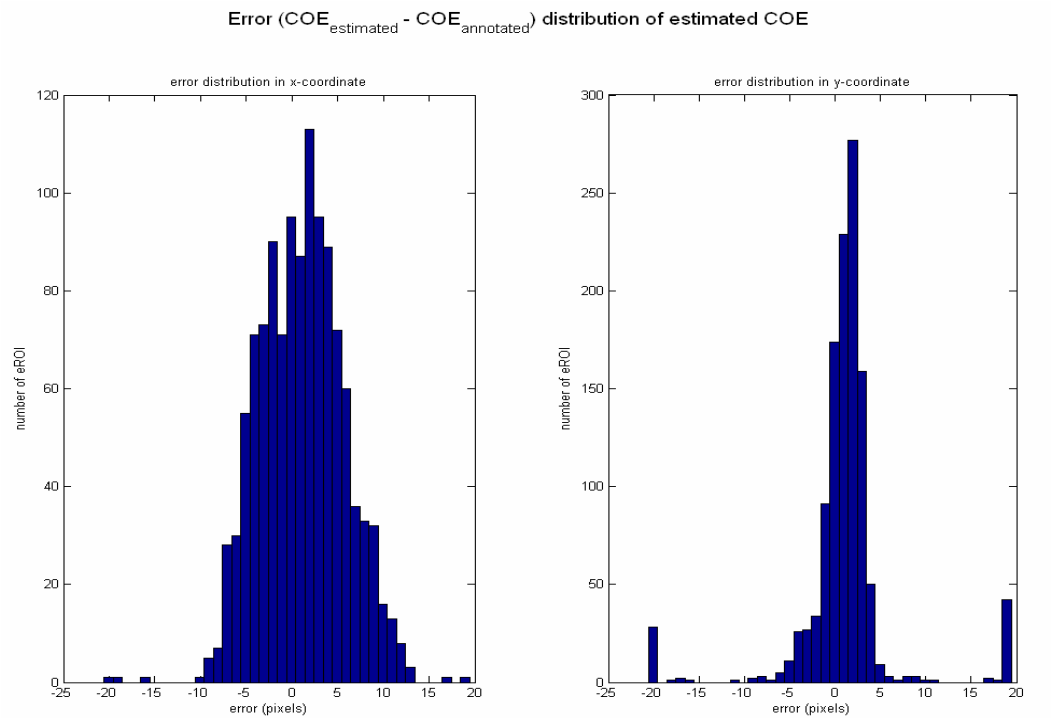


Figure 5-25. Histogram showing the error distribution in x and y coordinates of the estimated COE.

The SD of the estimated COE error distribution in both coordinates fall within the average size of an eye. Hence, the estimated COE position can be used to differentiate the position of an eye

from non-eye features within majority of eROIs. However, the y-coordinate of COE error distribution, as shown in Figure 5-25, had a substantial number of COEs estimated with large errors.

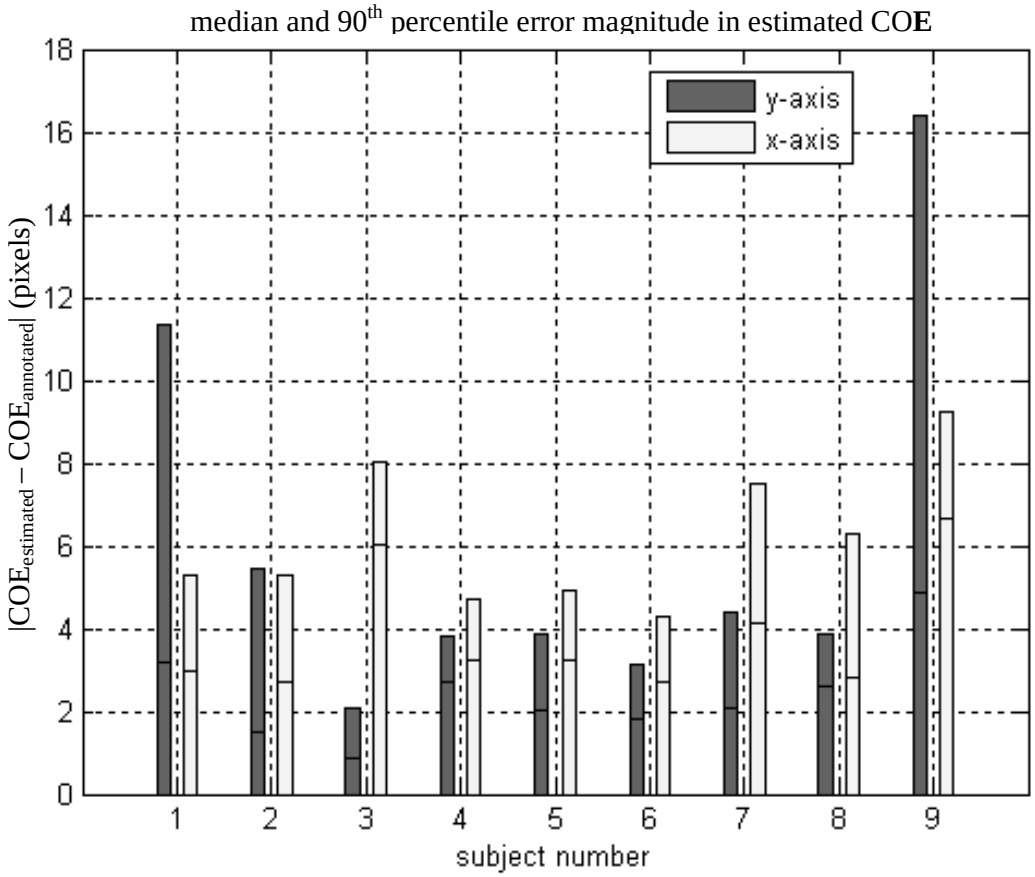


Figure 5-26. The median and 90th percentile COE detection error magnitude for the nine subjects in the reference database. The lower part of the bar graph represents the median error and the upper part represents the 90th percentile error in COE detection. The COE detection in y-axis was particularly poor for subject-1 and subject-9.

Table 5-4. The mean and the SD of medians and 90th percentiles of error magnitude in x and y coordinates of COE estimated for the N=9 subjects.

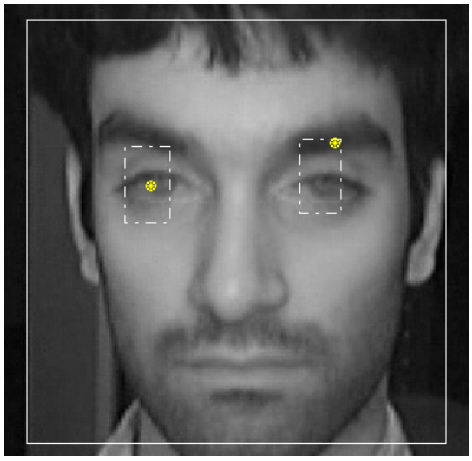
COE coordinates	mean ± SD of median error magnitudes (pixels)	mean ± SD of 90th percentile error magnitudes (pixels)
x	3.7 ± 6.4	5.9 ± 1.7
y	2.1 ± 1.4	5.8 ± 4.8

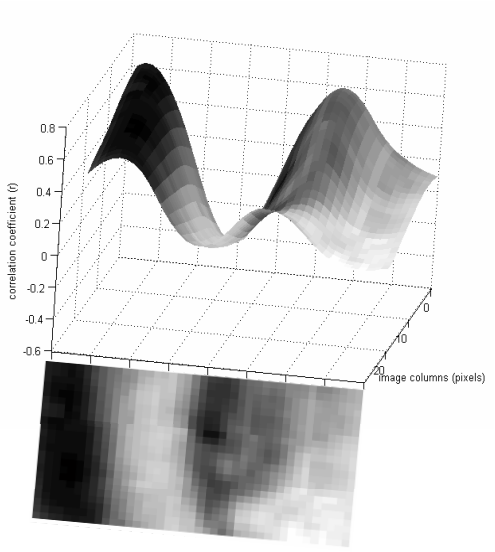
Further analysis of error magnitude for each individual subjects showed that the large errors in y-coordinates of some of the estimated COE were due to inaccurate detection of COE in two particular subjects. Figure 5-26 shows median and 90th percentile of the error magnitudes in x and y coordinates of estimated COE in the nine subjects.

Table 5-4 shows the mean and SD of these medians and the 90th percentiles of COE error magnitudes. The 90th percentile of COE error magnitudes in y-coordinate for subject 1 and 9 were significantly higher than other subjects, while the medians of COE error magnitudes for these two subjects were relatively similar to rest of the subjects. This result suggested the estimated COE error was large in some eROIs and relatively small in the majority of eROIs of these two subjects.

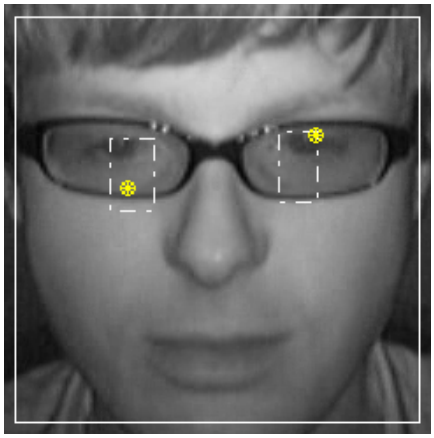
Frame-by-frame analysis of these two subjects showed that 15.2% (20 out of 132 eROI) eROI of subject 1 and 28.7% (38 out of 132 eROI) eROI of subject 9 had the COE's y-coordinate error magnitude greater than 10 pixels. As shown in left eROI of subject-1 in Figure 5-27(a), the main reason for large error magnitude in estimation of COE in subject 1 was due to false detection of dark and thick eyebrow as the true COE. As suggested by the corresponding **CC** of the left eROI in Figure 5-27(b), the false detection of the eyebrow was due to slightly higher correlation coefficient at the eyebrow region than the true COE. For subject 9, the main reason for large COE error magnitude was false detection of bottom part of the thick black glass-frame worn by the subject, as shown in right eROI in Figure 5-27(c) and Figure 5-27(d). This high correlation coefficient in the glass-frame region was due to the higher contrast in intensity between light skin region below the eye and dark glass-frame then the contrast in intensity between light skin regions surrounding the dark eye.

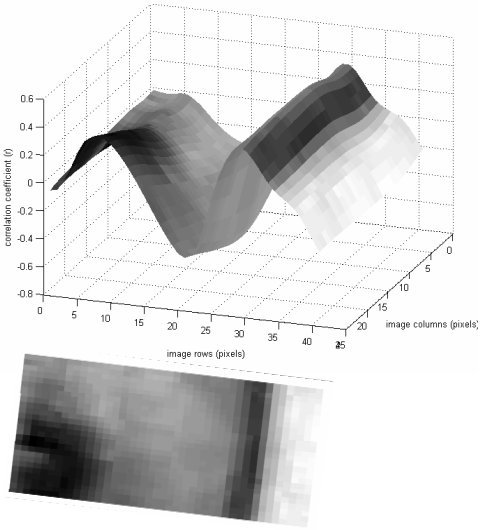
Since the eyelid detection method relied on accurate detection of COE (section 5.6.5), it was important to improve the COE detection method so that number of false detection of eyebrow and glass-frame could be reduced. In addition, eyebrow and glasses are common features in the eye region, so the COE detection method should be able to avoid these features for it to be practical. The next section discusses a method proposed to improve the COE detection method.





(a) (b)





(c) (d)

Figure 5-27. (a) Example showing false detection of eyebrow as the COE in left eROI in subject-1. (b) The 3D mesh plot of CC of the left eROI shows that the *r* at eyebrow is only slightly higher then at the true COE. (c) Image of subject-9 where the COE in the left eROI is falsely detected. (d) the 3D mesh plot of the CC matrix of the left eROI shows that correlation coefficient at the frame of the subject's glasses is much higher than the true COE.

5.6.4 Improving COE detection with Gaussian weighting

The performance of COE detection can be improved by weighting the **CC** matrix so that it emphasizes the region where the COE is most likely to be present and suppresses the regions of non-eye features within the eROI. The weighting function was derived by using prior knowledge about the distribution of annotated COE_a with respect to the centre of the eROI.

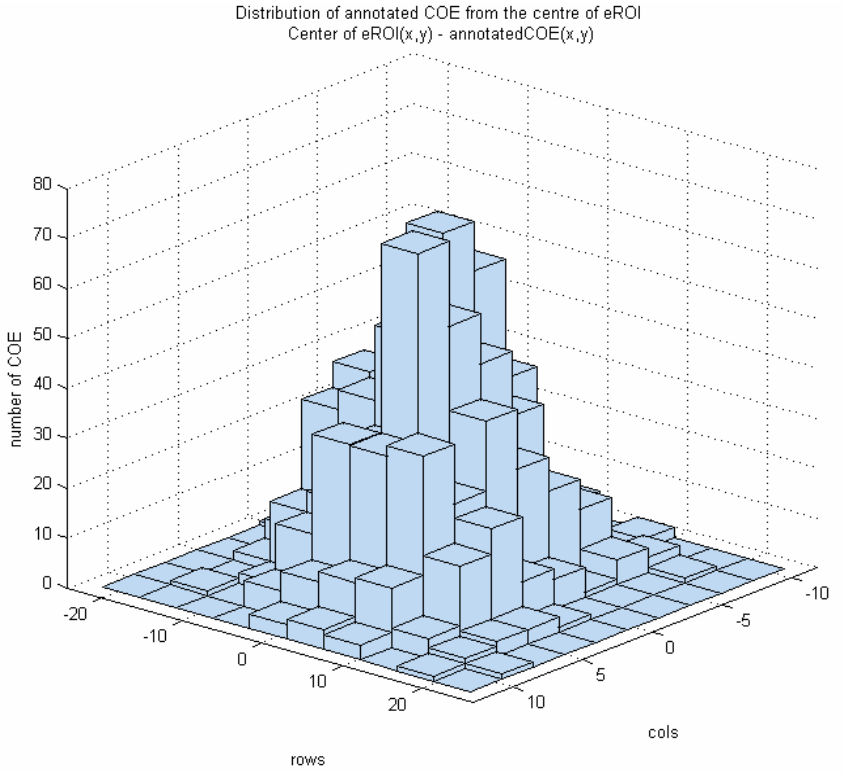


Figure 5-28. Histogram showing the position of the annotated COE in reference to the corresponding centre of eROI. This histogram show a Gaussian distribution of COE where most of the COE are positioned close to the centre of eROI and only few COE are found at the edges of the eROI.

Figure 5-28 shows a histogram of the distribution of the annotated COE from the centre of the eROI of both eyes in each frame in the annotated reference database. This histogram shows that the distribution has an approximately a Gaussian shape and can be represented by a 2D circularly symmetric Gaussian function **G**(*x*, *y*) as shown in Figure 5-29.

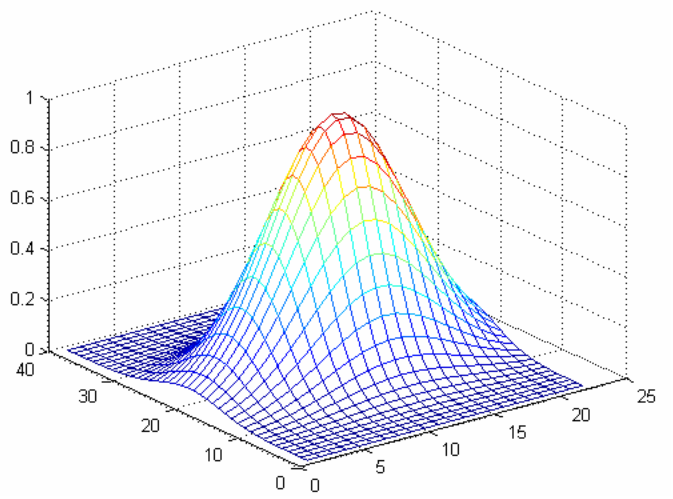


Figure 5-29. An example of a 2D Gaussian weighing function, G(x, y).

Hence, to improve the detection of COE, the **CC** matrix of each eROI is multiplied with Gaussian matrix **G**(*x*, *y*) as shown Equation 5-9, to form weighted correlation matrix $\mathbf{CC}_g$.

$$\mathbf{CC}_g = \mathbf{CC} \times \mathbf{G}(x, y)$$

Equation 5-9

Figure 5-30 shows an example where a falsely estimated COE in the left eROI of subject-1 was corrected by multiplying the **CC** of the eROI with a Gaussian matrix. In this eROI, the Gaussian matrix suppressed the false peak at the eyebrow region, which resulted in making the peak towards the true COE highest and correctly estimated as the COE position. In addition, the COE in the right eROI of the subject, which was already correctly estimated, was not affected by weighting the corresponding **CC**.

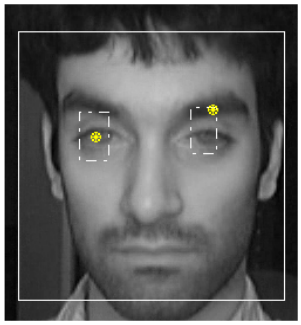
The respective left and right Gaussian functions shown in row two of Figure 5-30(b) and Figure 5-30(c) appear different because the parameters of the Gaussian matrix are dynamically derived for each eROI independently. For example, the size of the Gaussian matrix is set to be of same size as the **CC** of an eROI to be corrected.

Equation 5-10 gives the formula for deriving a Gaussian matrix. The centre (x_c, y_c) and standard deviations (σ_x, σ_y) of the Gaussian matrix are derived from statistical information about the distribution of annotated COE (COE_a) in the reference database. The centre coordinates (x_c, y_c) of Gaussian matrix is defined as the mean of respective x and y-coordinates of all COE_a. Separate centres were derived for the left and right eROIs. Since the Gaussian matrix is chosen to be circularly symmetric the σ_x and σ_y will be equal and is represented as single SD (σ). The σ of the Gaussian matrix is derived by calculating the mean SDs of COE_a distribution within both left and right eROI.

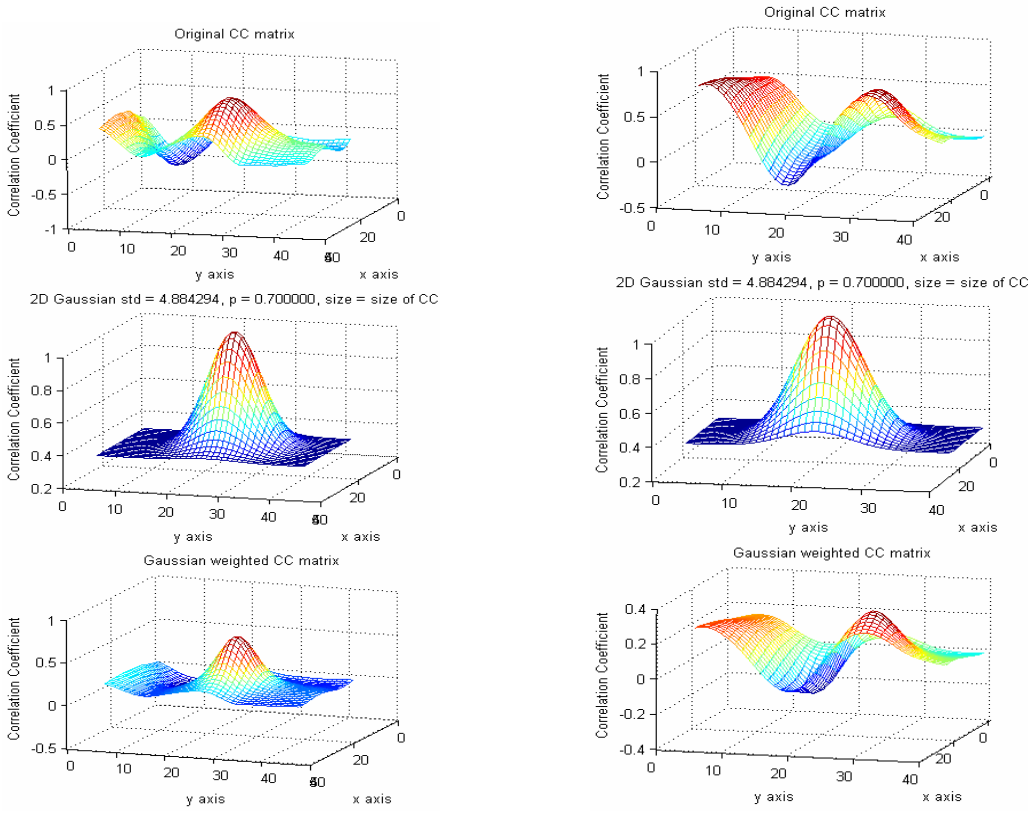
$$u(x, y) = \left(\frac{x - x_c}{\sigma_x}\right)^2 + \left(\frac{y - y_c}{\sigma_y}\right)^2$$

$$\mathbf{G}(x, y) = (1 - p) + \left(p \times \exp\left(\frac{-u(x, y)}{2}\right)\right)$$

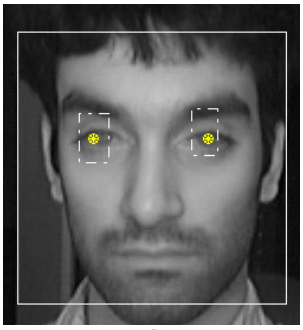
Equation 5-10



(a)



(b) (c)



(d)

Figure 5-30. (a) Image showing the false estimation of eyebrow as the COE in the right eROI and correct estimation of COE in the left eROI. (b) Three rows sequentially showing the mesh plot of raw CC, Gaussian weighting function, and the weighted CC for the left eROI. (c) Same plot as (b) for the right eROI, where peak at the eyebrow region in CC is suppressed due to Gaussian function weighting making the peak close to the true COE towards the centre of the eROI highest peak in the CC. (d) Image showing correct COE estimation for the right eROI and unaffected COE estimation for the left eROI after the correction of COE estimation with Gaussian weighting function.

The Gaussian matrix is normalized so that its peak value is set to 1 and its minimum value is set to $(1 - p)$. In Equation 5-10, p is a scaling factor, which defines the lower weighting limit of a Gaussian matrix. In other words, the p value of Gaussian matrix affects the weight put on edges of **CC** matrix. To determine an optimum p value for general population in the collected reference database, the **CC** matrix of each eROI in annotated frames was multiplied with different Gaussian matrixes derived with various p values ranging from 0 to 1 with increment of 0.1. Then, the p value for which the COE detection algorithm produced the most accurate (lowest error magnitude) result was selected as the optimum fixed p value used for every Gaussian matrix.

Figure 5-31 shows means of 90th percentile error magnitudes (plot on the left) and means of median error magnitudes (plot on the right) in estimating the COE's y-coordinate of nine subjects based on various $\mathbf{CC}_g$ derived with Gaussian matrixes with various p values. Each plot in Figure 5-31 shows the error magnitude in COE detection for three sub-groups of subjects, which are five subjects without glasses, four subjects with glasses, and combination of all nine subjects.

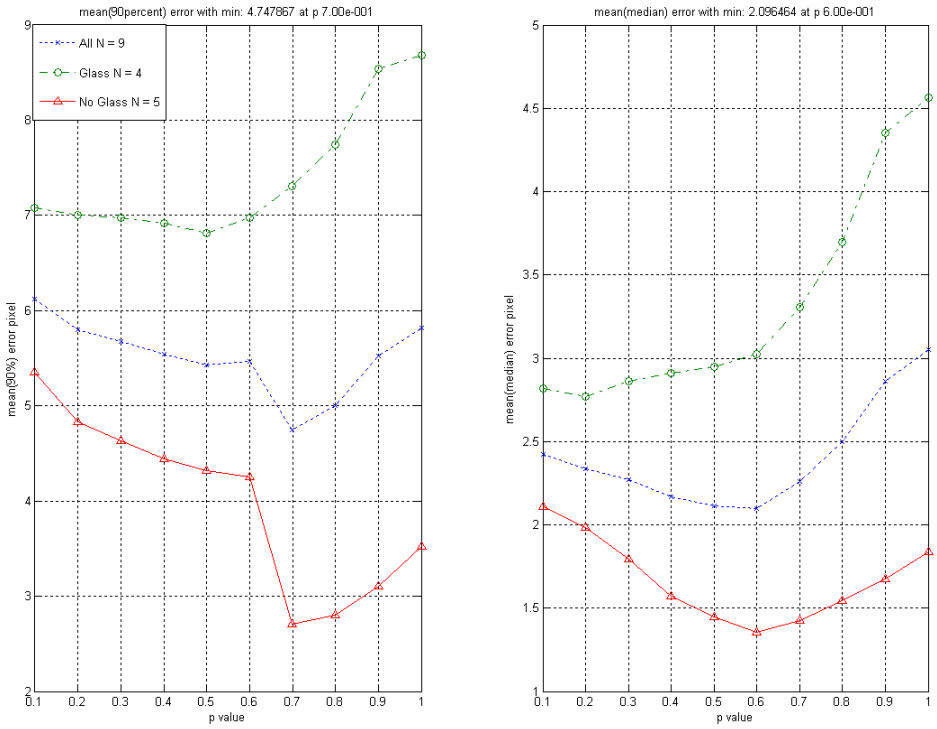


Figure 5-31. Left: mean 90th percentiles; right: mean medians; of errors in detection of y-axis of COE.

In the mean 90th percentile error magnitude plot of all nine subjects, the COE detection algorithm had the lowest error magnitude of 4.7 pixels in estimating COE when the *p* of the Gaussian matrix was set to 0.7. The mean medians error magnitude plot of the nine subjects indicated a lowest error magnitude of 2.1 pixels in estimating COE when the *p* was set to 0.6. The optimum *p* value for the Gaussian matrix was set to 0.7 as indicated by the 90th percentile plot because it represents the worst-case scenario. In addition, the mean median performance of COE detection is not compromised by setting the *p* value to 0.7 because there is very small difference between the average median error magnitudes when $p = 0.6$ and $p = 0.7$.

5.6.5 Performance of improved COE detection

The Gaussian correction of the **CC** reduced the overall COE estimation error. The mean and SD of the error distribution in both x and y coordinates of the COE was (mean ± SD) 0.3 ± 4.9 pixels and 1.3 ± 5.0 pixels, which is a slight improvement compared to initial performance presented in section 5.6.2. Since the COE detection method was mainly developed to distinguish the vertical position of an eye from other facial features within an eROI, only the performance of the method to detect the COE's y-coordinate was further analysed in detail. Comparison of the corresponding error distribution plots of y-coordinate (right plot) in Figure 5-25 and Figure 5-32 show that the number of eROIs with large error both above and below the true COE was reduced.

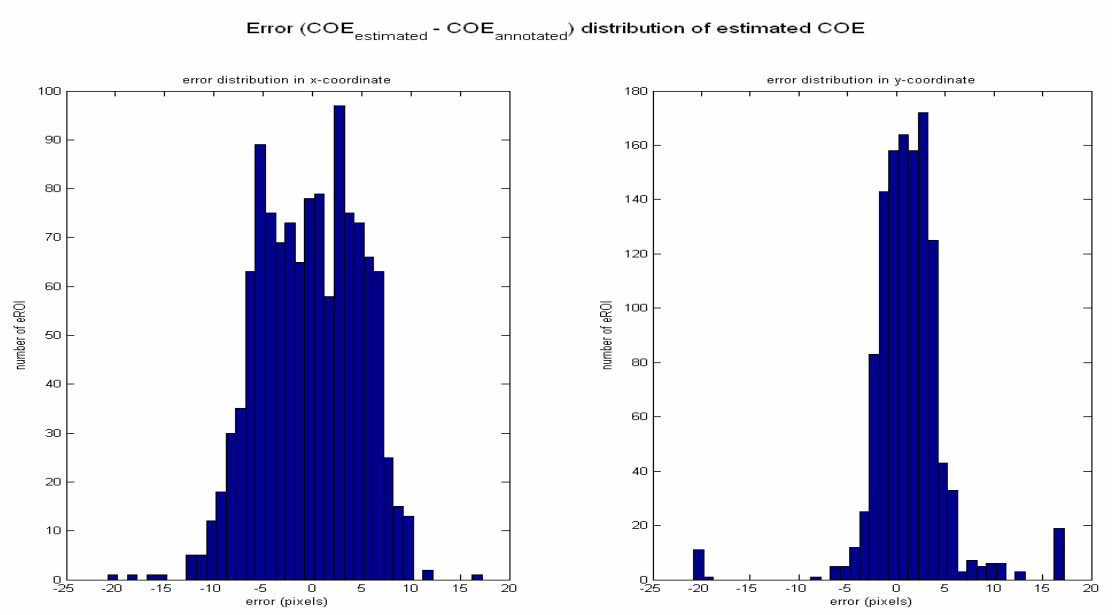
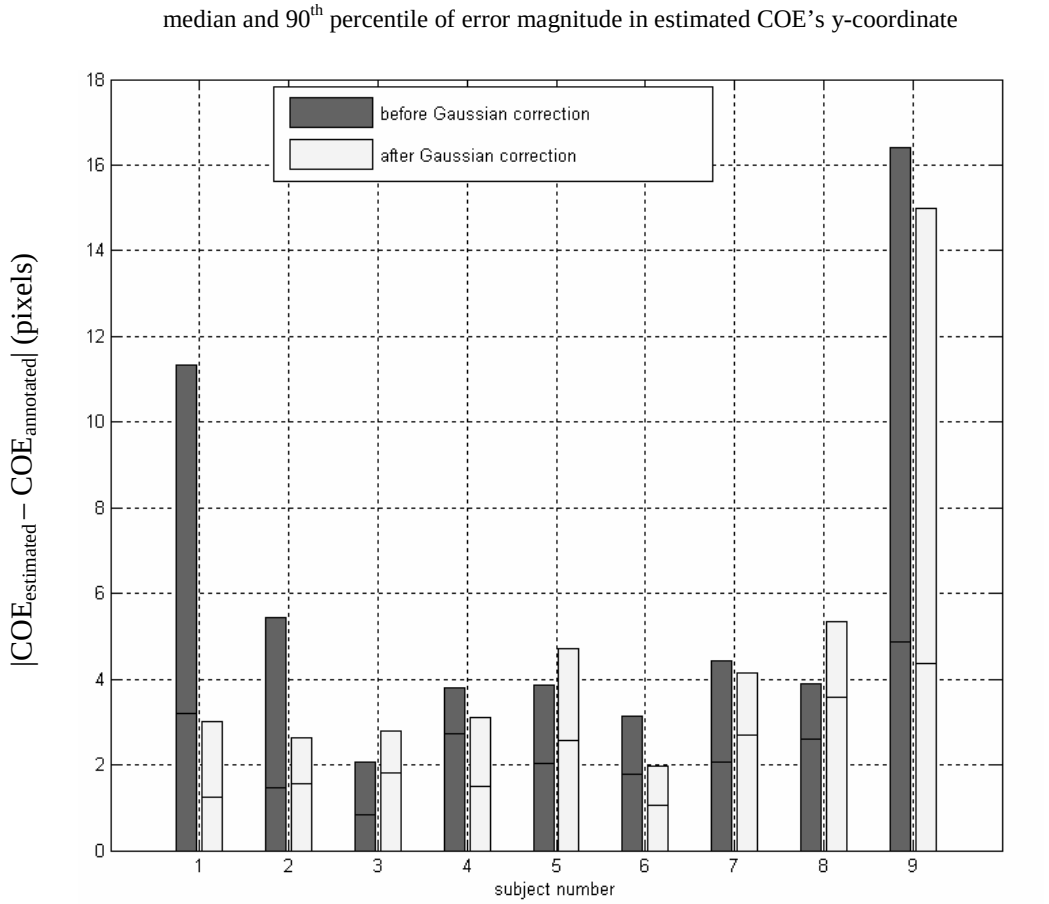


Figure 5-32. Histogram showing the error distribution in x and y coordinates of the estimated COE after applying Gaussian correction to the CC matrix.



Subject #	1	2	3	4	5	6	7	8	9
Before	11.3	5.4	2.0	3.8	3.8	3.1	4.4	3.8	16.4
After	3.0	2.6	2.8*	3.1	4.7*	1.9	4.1	5.3*	15.0

Figure 5-33. The 90th percentile error in detection of y-coordinate of COE before and after the correction of CC with the Gaussian matrix for all nine subjects; * slight increase in detection error.

Figure 5-33 shows that for subject 1 the Gaussian correction of **CC** considerably reduced the 90th percentile error magnitude in estimated COE y-coordinate from 11.3 pixels to 3.0 pixels. However, for rest of the subjects there was no substantial effect on error magnitude in y-coordinate of COE was observed. There was slight reduction in the error magnitude for subjects 2, 4, 6, 7, and 9. Particularly, for subject 2 and 6, the 90th percentiles of error magnitude were reduced by approximately 50%. On the other hand, there was slight increase in error magnitude in COE y-coordinate for subject 3, 5, and 8. The substantial improvement in COE detection for subject 1 due the Gaussian correction of **CC** was an encouraging result. However, the minimal improvement in subject 9 was unsatisfactory as both subject 1 and 9 contributed to overall performance degradation of COE detection method in initial analysis (section 5.6.2).

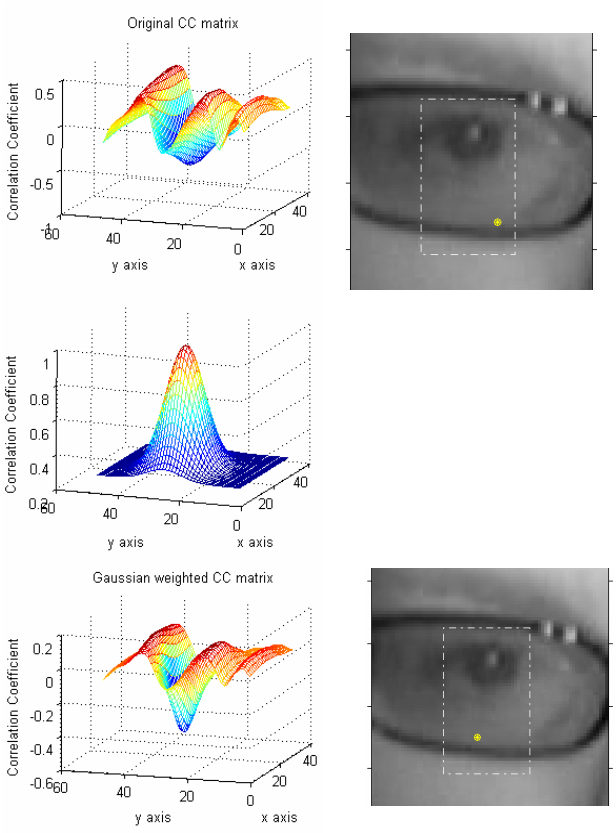


Figure 5-34. Example showing no improvement in COE y-coordinates estimation for subject 9 even after the Gaussian correction of CC because both eye and glass-frames lie close towards the edge of the eROI. top row: before Gaussian correction; middle row: Gaussian matrix used; bottom row: After Gaussian correction.

The Gaussian correction of **CC** did not reduce the 90th percentile error magnitude of COE y-coordinate estimation in subject 9 because in most eROI with large estimation error, eye and glass-frames were positioned at the opposite edges of eROI as shown in Figure 5-34. Since the Gaussian matrix weighs opposite edge of **CC** equally, the relative peak *r* values at eye and lower glass-frame regions were not affected. So correcting the **CC** with a Gaussian matrix was not able to avoid false detection of non-eye features if the eye and falsely detected non-eye features lie at the opposite edges of eROI. Therefore, the reference data of subject 9 was not used in evaluation of eye-feature detection methods presented from this point onwards. In addition, reference data of subject 8 was also discarded in performance evaluation of further eye-feature detection methods developed because of the poor quality of annotated image data for the subject. In a large number of annotated frames, glare from reflection of NIR source on the glasses worn by subject 8 covered the large part of eye making parts of the eye invisible.

Hence, the performances of methods developed in this project were evaluated based on subjects 1 to 7.

Table 5-5: Performance of the developed COE y-coordinate detection method based on 7 subjects.

Group	Number of subjects (N)	mean ± SD of median error magnitudes (pixels)	mean ± SD of 90th percentile error magnitudes (pixels)
Subjects with glasses	2	2.6 ± 0.1	4.4 ± 0.4
Subjects without glasses	5	1.4 ± 0.3	2.7 ± 0.4
All subjects	7	1.7 ± 0.6	3.2 ± 0.9

Table 5-5 shows the performance of the developed COE detection method for detecting COE's y-coordinate after applying it to 924 eROIs (462 frames) of 7 subjects. Table 5-5 lists the mean and SD of median and 90th percentile of error magnitudes in estimating the y-coordinate of COE. Between the two groups of subjects with and without glasses, the COE detection method performed much better for the group without glasses. The averages of error magnitudes from both of these groups were taken to derive the overall performance of the developed COE detection method. On average, the developed COE detection method estimated the y-coordinate of COE within 3.2 pixels with SD of 0.94 pixels in 90% of the eROIs. Additionally, the method estimated the y-coordinate of COE within 1.7 pixels on average with SD of 0.64 pixels in median number of eROIs analysed.

The overall mean 90th percentile error magnitude of 3.2 pixels for estimation of the y-coordinate of COE was considered sufficient to distinguish the vertical position of an eye from other non-eye features within an eROI. The average number of pixels between vertical distance of eye and eyebrow of the subjects in the reference database was greater than 15 pixels. Once the COE was detected within the eROI, the positions of eyelids were searched using the eyelid detection method as discussed in the next section.

5.7 Eyelid detection

The upper eyelid (UEL_y) and lower eyelid (LEL_y) must be detected to determine the degree of eye closure (see sections 1.9 and 5.1). In this report, unless specifically specified otherwise, the UEL_y and LEL_y refer to the y-coordinates of apex of the upper and lower eyelids, respectively. The apex of an eyelid is the point in the eyelid margin furthest vertically from the eye centre, and the eyelid margin is the parabolic inner edge of the eyelid (see section 1.3 for basic anatomical description of the eye). Hence, estimating the UEL_y and LEL_y will allow eye closure to be measured by calculating the vertical distance between the two apices. The UEL_y and LEL_y positions were detected by identifying change in mean image intensity at corresponding edges of the eye. However, before the positions of the eyelids were detected, the eROI was redefined to optimize it for eyelid detection.

5.7.1 ROI for eyelid detection

The eROI used for COE detection is not ideal for eyelid detection as it was derived from distribution of annotated COE_a (see section 5.4.1) and may exclude eyelids under some circumstances. For example, if the position of COE was detected near the top or bottom edge of an eROI, it was highly likely that one of the eyelids will be excluded from the eROI making them impossible to be detected. Hence, a new eyelid ROI (elROI) was defined before detecting the eyelids.

The position and dimension of the elROI were defined so that both eyelids were highly likely to be present within the elROI. The centre of elROI was set equal to the estimated COE position because the y-coordinate of the COE (COE_y) vertically lay between UEL_y and LEL_y and its x-coordinate aligns with the horizontal position of apex of the eyelids. Figure 5-35 shows an example where the estimated COE (marked with '*****') within the eROI (marked with white dashed line) was used to defined the centre position of elROI (marked with dark solid line).

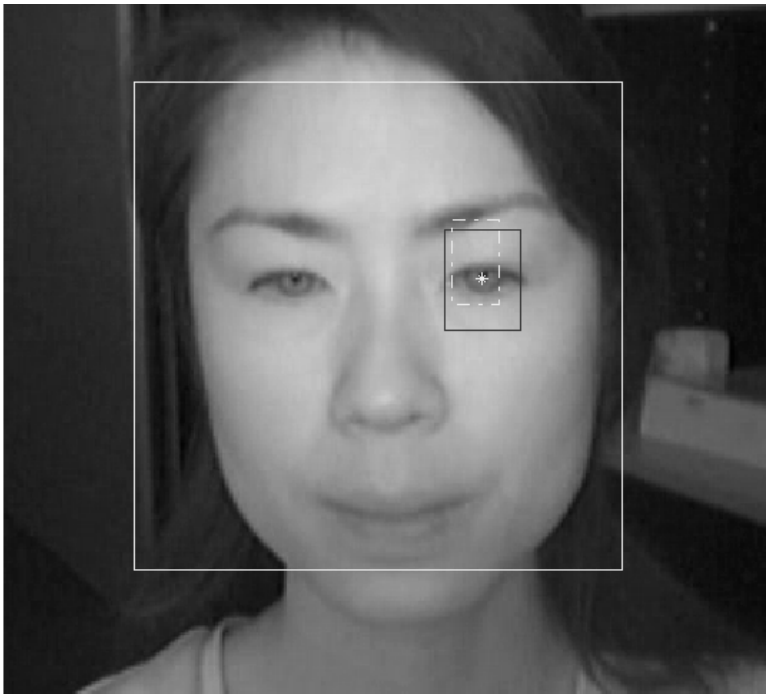


Figure 5-35. The inner ROI marked with dark solid line defines the elROI for eyelid detection. It is centred at the estimated COE position marked with '*'. The inner dashed line defines the eROI used for COE detection and the outer ROI marked by solid white line defines the fROI.

Since the height of the eye template image used for COE detection represents the average height of an open eye in the reference database (see section 5.6.1), the height of elROI was assigned to be twice the height of the eye template image. Making the elROI twice as high as an average open eye ensures the presence of both upper and lower eyelids within the elROI, even if the COE is estimated at the edge of the eROI. It should be noted that since the size of the eye template image was not scaled in proportion to the fROI, the height of the elROI also remained constant in all analysed frames. Using elROI with a constant height is not an ideal approach as the size of the fROI can scale within an image depending on the distance of the subject from the camera. However, the inter-subject variability in facial size was assumed to be small in this project because the subjects were seated at a fixed distance from the camera during the collection of the reference video data. Hence, keeping the height of the elROI constant would have minimal effect on the overall result in this project.

Initially, the width of elROI was set to be same as the width of eye template image. The result of this initial trial showed that the performance of eyelid detection method was particularly poor in subjects who wore eye-glasses because parts of eye-glasses were being falsely detected as the eyelids. In addition, the horizontal position of elROI was unreliable because the accuracy

of the estimated COE in x coordinate was poor (see section 5.6.2). Hence, to find an optimum width of the elROI, the developed eyelid detection method was applied to elROIs with various widths and a width of elROI for which the method had the best performance was assigned as the fixed width of elROI. The optimum width of elROI was set to 37 pixels based on the results of the eyelid detection method discussed later in section 5.7.4.

5.7.2 Vertical integral projection

The difference in average image intensity between the dark eye region and its surrounding light skin region was used for detecting the UEL_y and LEL_y. Figure 5-36 shows an elROI with manually annotated positions of UEL_y, COE_y (y-coordinate of COE), and LEL_y. From top to bottom, the average image intensity visually transitions from light skin region to darker eye region at UEL_y position and the average image intensity transitions from dark eye region to lighter skin region at LEL_y position. Hence, the corresponding positions at the edges of the eye where the transition in mean image intensity takes place were assigned as the UEL_y and LEL_y.

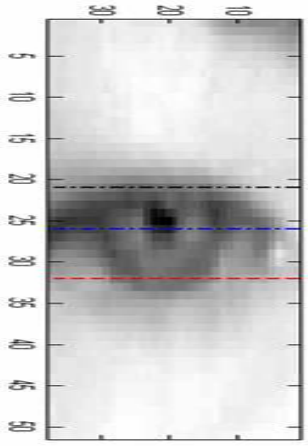


Figure 5-36. An elROI with the manually annotated vertical positions of UEL_y, COE_y, and LEL_y.

The vertical integral projection (*VIP*) of an image captures change in mean image intensity along the rows of the image. *VIP* at nth row of a grayscale elROI sub-image of N x M resolution was derived by calculating the mean intensity in each row as

$$VIP(n) = \frac{\sum_{m=1}^{M} I_{elROI}(n,m)}{M}$$

Equation 5-11

Gradient of *VIP* (*VIP'*) was calculated using a derivative function based on linear regression as shown in Equation 5-12. Given a set of *N* equi-spaced samples, *VIP*(0), *VIP*(1), … *VIP*(*n*), … *VIP*(*N*-1), the gradient of the function at sample *VIP*(*n*) was estimated based on the slope of the line of best fit through *k* consecutive samples centered at *VIP*(*n*). For simplicity, *k* was chosen to be odd number. The unit for the gradient is change in *VIP* per sample.

$$VIP'(n) = \frac{\sum_{s=\frac{k-1}{2}}^{\frac{k-1}{2}} sVIP(n+s)}{\sum_{s=\frac{k-1}{2}}^{\frac{k-1}{2}} s^2}, \qquad n = \frac{k-1}{2}, \ldots, N - \frac{k-1}{2}$$

Equation 5-12

In Equation 5-12, the gradient at (*k*-1)/2 samples at the start and end of *VIP* are not estimated. Since it is highly unlikely for substantial change in mean image intensity at the edge of elROI, the gradient at these start and end samples was assigned to a constant value as shown below.

$$VIP'(n) = VIP'\left(\frac{k-1}{2}\right) \qquad n = 1, \ldots, \frac{k-3}{2}$$

$$VIP'(n) = VIP'\left(N - \frac{k+1}{2}\right) \qquad n = N - \frac{k-1}{2}, \ldots, N$$

This method of gradient calculation is only suitable for data over-sampled with respect to the Nyquist limit, or where an accurate estimate is required only on relatively smooth regions. The size of the *k* sampling window in Equation 5-12 determines the smoothness of *VIP'*. Smoothing or low-pass filtering of the *VIP'* is required to remove high-pass noise that are produced by sudden change in image intensity in image artifacts such as eye glints, shadows, and reflections off the eye-glasses. The size of *k* sampling window was set to 11 samples based on experimental trial-and-error to produce optimum low-pass filter with minimal loss of original information.

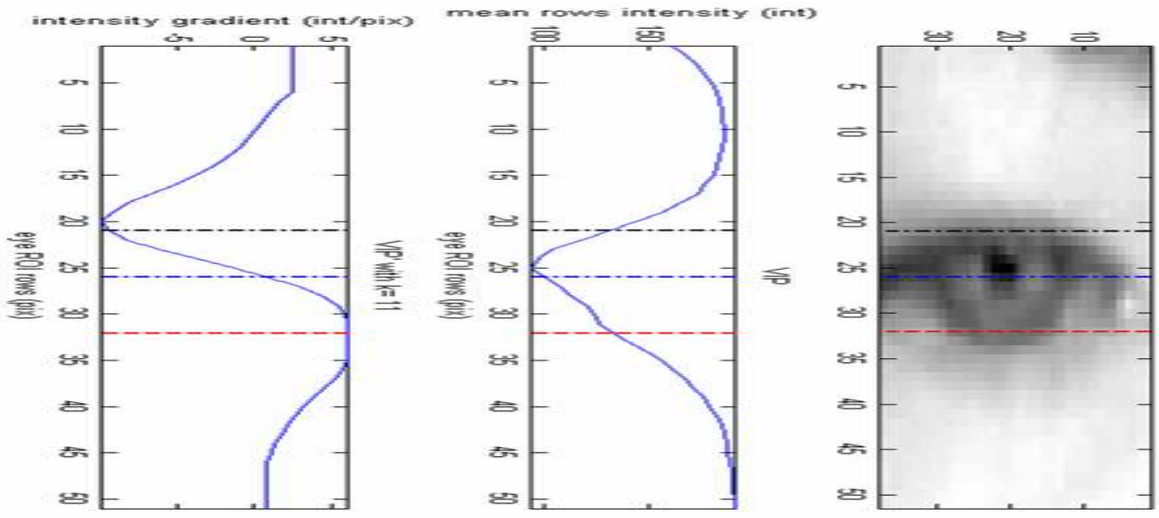


Figure 5-37. Right: elROI; Middle: *VIP* of the elROI; Left: *VIP'*, the gradient of *VIP*. In this figure, the manually annotated row position of apex of the upper eyelid and lower eyelid respectively aligns closely with the first local minima above the COE_y and first local maxima below the COE_y in *VIP'* of the elROI.

Figure 5-37 shows the *VIP* and *VIP'* plots of the elROI with lines marking the row positions of manually annotated UEL_y, COE_y, and LEL_y. In this figure, the UEL_y closely aligns with the first local minimum above the COE_y in the *VIP'*, which corresponds to the steepest negative slope above the COE_y in *VIP*. In contrast, the LEL_y aligns with the first local maximum below the COE_y in the *VIP'*, which corresponds to the steepest positive slope below the COE_y in *VIP*. Hence, the detection of the first local minimum above the COE_y and the first local maximum below the COE_y in *VIP'* of elROI can be used to estimate the respective UEL_y and LEL_y. To assess the reliability of these indicators of UEL_y and UEL_y positions in *VIP'* plot of an elROI, the *VIP* and *VIP'* plots of 80 randomly selected elROI from annotated frames in the reference database that included all 9 subjects were visual inspected.

5.7.3 Upper eyelid detection

Visual inspection showed that in *VIP'* plots for the majority of elROIs, the annotated UEL_y closely aligned with the first local minimum above the COE_y. Hence, the position of first local minimum above the estimated COE_y in *VIP'* of an elROI was assigned as the UEL_y (Figure 5-36). If there are no local minima above the COE_y, the first row of the elROI was assigned as the UEL_y.

5.7.4 Lower eyelid detection

Visual inspection of *VIP'* plot of the randomly selected eighty elROIs showed that annotated LEL_y did not consistently align with the row position of first local maximum below the annotated COE_y. A common reason for misalignment of LEL_y with the first local maximum below the COE_y was due to complex mean image intensity patterns below the COE_y. For example, in Figure 5-38, the mean image intensity below the COE_y changes from dark pupil to light region of iris, then to dark limbus ring (section 1.3) and finally to light skin region below the LEL_y. This contrast in intensity between features within an iris was further enhanced under infrared illumination (section 1.3). This change in mean image intensity pattern within the iris creates a local maximum above the lower eyelid edge of an eye. For example, in *VIP'* of the elROI in Figure 5-38, the first local maximum below the COE_y was formed at the edge of pupil and iris.

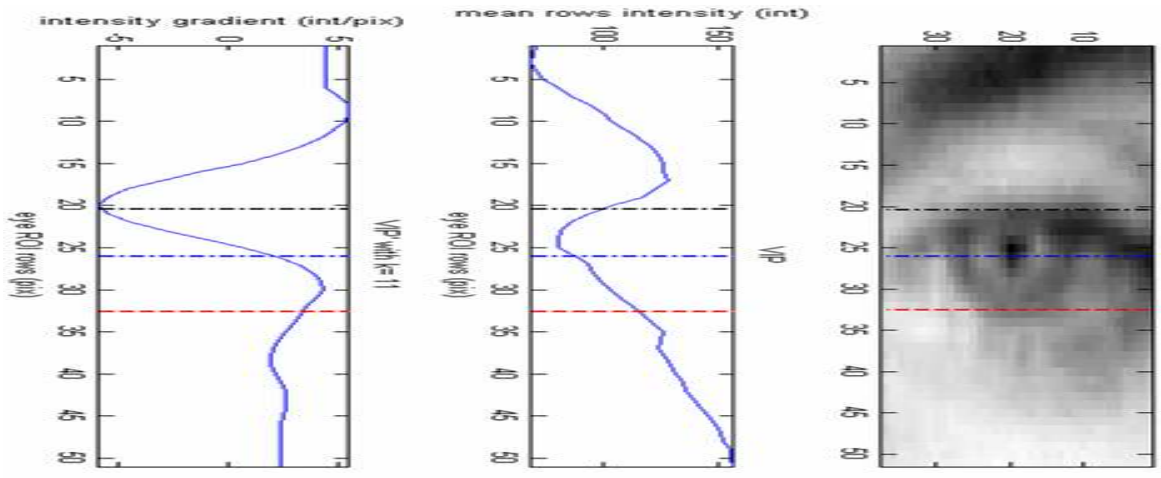


Figure 5-38. A figure illustrating that the local maxima below the COE_y in *VIP'* does not sometime align with annotated LEL_y as expected due to changes in mean image intensity between features within an iris.

Figure 5-39 shows an example of another condition under which the annotated LEL_y did not align with the first local maximum below COE_y in *VIP'*. When the subject gazed upwards, the lower portion of the iris and the sclera was exposed within the palpebrae fissure, resulting in the first local maximum below the COE_y in *VIP'* to be formed at the edge of the iris and sclera due to sudden change in mean image intensity.

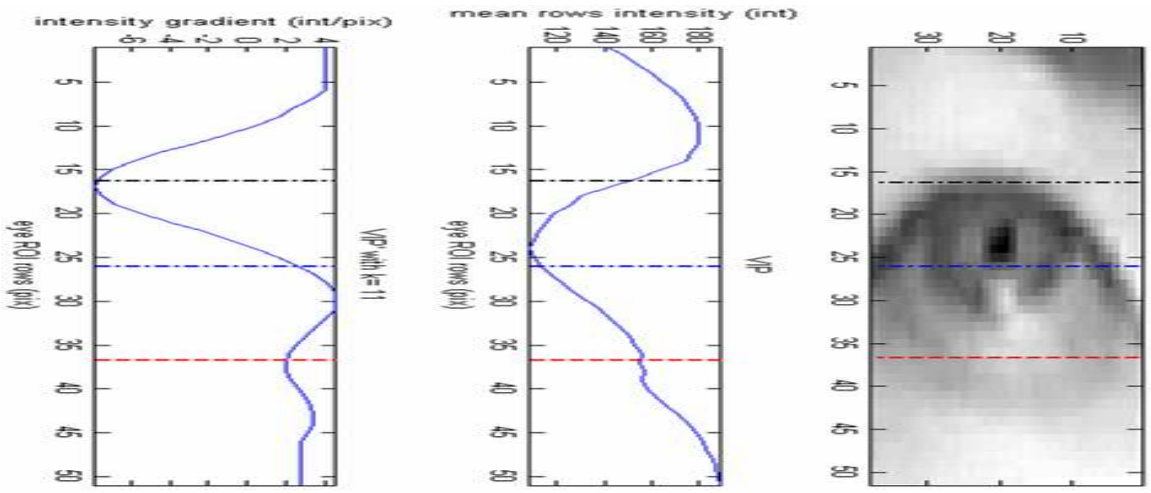


Figure 5-39. During upward gaze the first local maxima below the COE_y in *VIP'* is usually formed at the edge of iris and sclera instead of at the LEL_y.

As an alternative approach, the position of the furthest local maximum below the COE_y in *VIP'* was trialled as indicating LEL_y, because the lower eyelid is often the last eye feature with substantial contrast in mean image intensity from dark to light region. However, this alternative indicator of LEL_y did not improve the accuracy of LEL_y detection. Hence, the first local maximum below the COE_y was used as the indicator of LEL_y position although it was considered to be a less reliable indicator of LEL_y. If there are no local maxima below the COE_y, the last row of elROI is assigned as the LEL_y.

5.7.5 Performance of eyelid detection

Figure 5-40 shows an example of performance of the eyelid detection method. This figure shows the same elROI and its *VIP* and *VIP'* plots as in Figure 5-37, with additional solid lines indicating the positions of the estimate UEL_y, COE_y, and LEL_y. In this example, the error magnitude of both the estimated UEL_y and LEL_y was within one pixel. The overall performance of the eyelids detection method was quantitatively evaluated by analysing the error and error magnitude in the estimated eyelid positions. The performance of the upper and lower eyelid detection are discussed separately in next two sections.

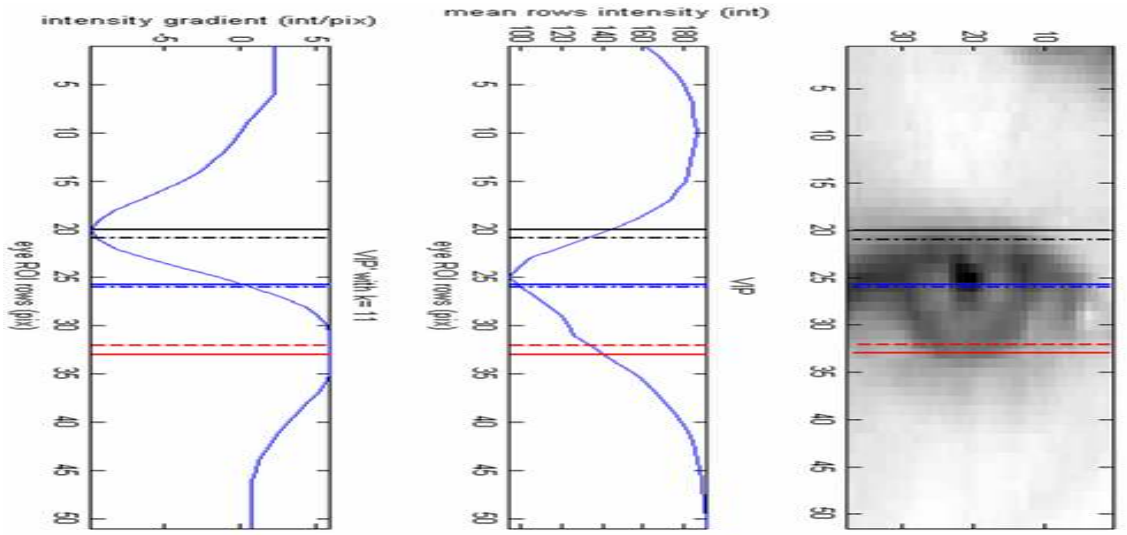


Figure 5-40. Figure showing both the estimated (solid line) and annotated (dotted line) row positions of the UEL_y, COE_y, and LEL_y.

5.7.5.1 Upper eyelid

Apart from the annotated frames where the subjects had their eyes fully closed, the error in estimated UEL_y was calculated by taking the difference between the estimated UEL_y and the y-coordinate of top edge of the annotated visible eye ROI (refer to section 4.2.2 for annotation of eye features). In frames with fully closed eyes, the error was calculated by taking the difference between the estimated UEL_y and the y-coordinate of bottom edge of the annotated visible eye ROI because it was assumed that when the eyes are fully closed the apex of both eyelids would overlap at the same position.

As discussed in section 5.7.1, results of eyelid detection under various widths of elROI were analysed to select an optimum width of elROI. For this purpose, the mean of 90^{th} percentile error magnitude and mean of median error magnitude of the estimated UEL_y from the 7 subjects were calculated and analysed for each result of applying the eyelid detection method to elROIs of various widths. The width of elROI was varied from 5 to 161 pixels with the interval of 4 pixels. Figure 5-41 show these means of 90^{th} percentile and median error magnitude of the estimated UEL_y for each elROI. The result of UEL_y estimation is also further divided into a group with two subjects who wore eye-glasses and another group of five subjects who did not wear eye-glasses.

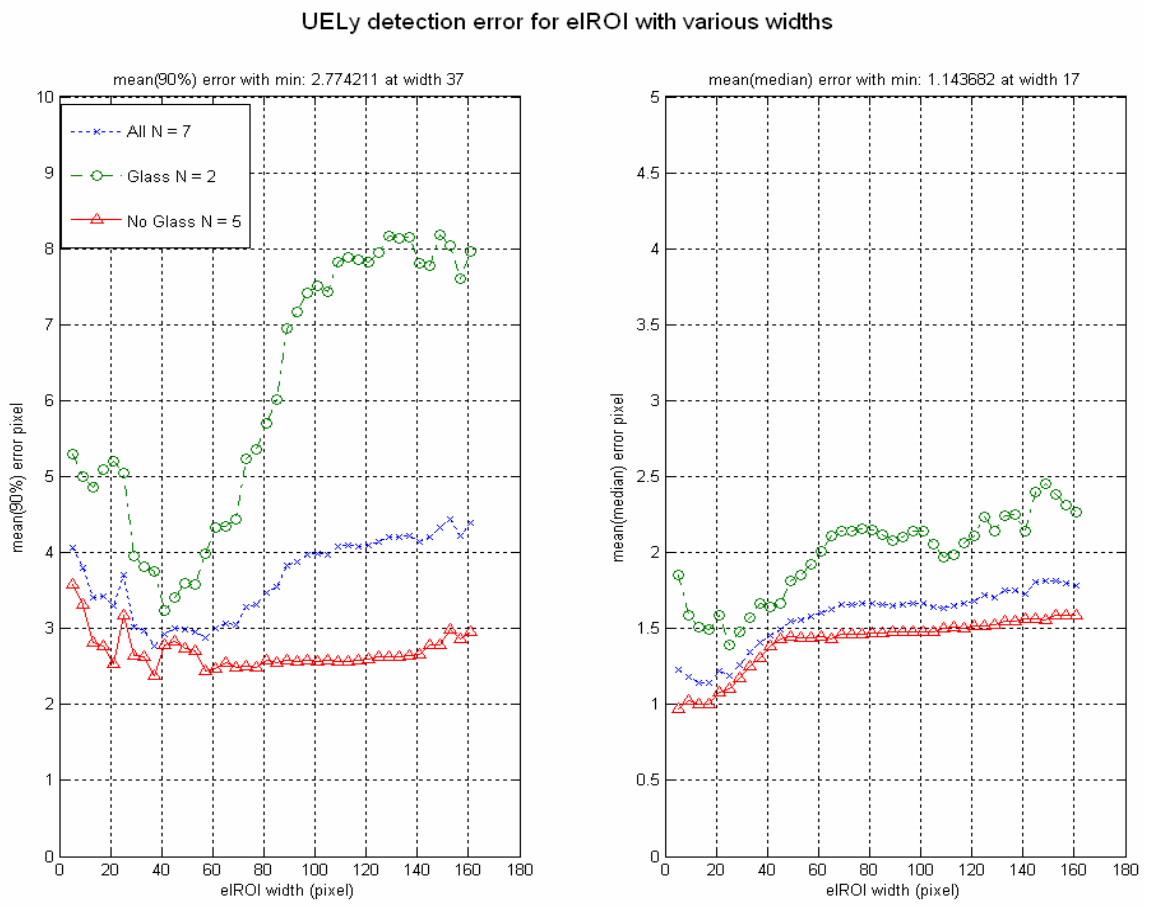


Figure 5-41. Means of 90th percentile and median error magnitude of UEL_y estimation under various widths of elROI.

Based on the mean of 90th percentile error magnitude plot in Figure 5-41, the eyelid detection method has the best performance in UEL_y estimation, with only 2.7 pixels when the width of elROI is set to 37 pixels. The lowest mean of median error magnitude of the estimated UEL_y was only 1.1 pixels when the width of elROI was set to 17 pixels. Since the 90th percentile error magnitude represents a measure of reliable worst-case performance of a method, the optimum width of elROI was set to 37 pixels. Therefore, all the results presented here onwards are based on the width of elROI set to 37 pixels. In real-world environment, use of a constant width of elROI in video of all subjects is not an ideal approach as the size of a face can vary in video due to change in distance between the camera and subject. In addition, slight inter-subject variability in facial size also exits. Hence, it would be better to assign the width of elROI as a proportional parameter of width of fROI because the Haar-face detection algorithm can detect a face under varying scale. However, use of a constant width of elROI does not affect the

overall performance of eyelid detection in this project because the distance between camera and subjects was kept constant during the collection of the reference video data.

As listed in Table 5-6, when the width of elROI was set to 37 pixels, the mean of 90th percentiles and mean of medians error magnitude of the estimated UEL_y from the seven subjects was 2.7 pixels and 1.4 pixels, respectively. As expected, the UEL_y estimation was more accurate for subjects without glasses, as image artifacts from glasses can confuse the eyelid detection algorithm. The mean 90th percentile error magnitude of only 2.7 pixels for UEL_y estimation is an encouraging result for using it to measure eye closure. An average height of a visible eye ROI (distance between apex of upper and lower eyelid) in the reference database was 14.2 pixels when the eyes are fully open. So theoretically, based on the performance of the estimated UEL_y and assuming a perfect LEL_y estimation, the eye closure should be able to be measured within 90.1% of the true measure in general and within 81% of the true measure in worst-case scenario.

Table 5-6. The means and SDs of the 90th percentile and median error magnitude of the estimated UEL_y in the 7 subjects with the width of elROI set to 37 pixels.

Group	Number of subjects (N)	mean ± SD of median error magnitude in UEL_y (pixels)	mean ± SD of 90th percentile error magnitude in UEL_y (pixels)
Subjects with glasses	2	1.6 ± 1.8	3.7 ± 4.1
Subjects without glasses	5	1.3 ± 2.0	2.4 ± 3.2
All subjects	7	1.4 ± 2.0	2.7 ± 4.2

Figure 5-42 shows the error distribution of the estimated UEL_y in elROIs in annotated frames. Mean of estimated UEL_y error distribution showed a small bias of 0.3 pixels below the true UEL_y with the SD of 4.2 pixels. This error distribution result of the estimated UEL_y further show that in general the estimation of the UEL_y was good in majority of elROI.

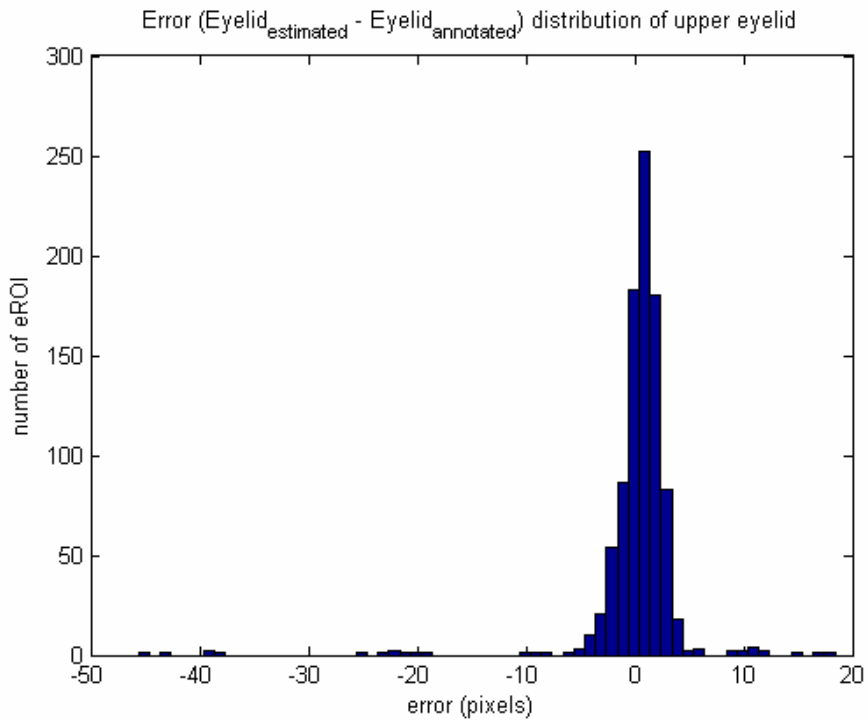


Figure 5-42. Error distribution of the estimated upper eyelid position.

5.7.5.2 Lower eyelid

Figure 5-43 shows the error distribution of the estimated LEL_y positions in annotated elROIs. The mean of the error distribution indicated a slight bias of 0.38 pixels below the true LEL_y with SD of 4.9 pixels.

In large number of elROI, the estimated LEL_y errors were within the range of -2 pixels (above true LEL_y) to +4 pixels (below true LEL_y). Further analysis of the eye images with LEL_y estimation error greater than 5 pixels showed that the false detection of LEL_y below the true LEL_y were mainly found in elROI with either fully closed or ¾ closed eyes. In an elROI with a closed eye, as shown in Figure 5-44, the LEL_y is estimated to be at the bottom of the eyelashes. However, in frames with fully closed eyes the LEL_y was manually annotated to be at the top of the eyelashes (section 4.2.2). Hence, this difference between the estimated and annotated LEL_y positions resulted in large positive LEL_y estimation error in elROI with either fully or ¾ closed eyes.

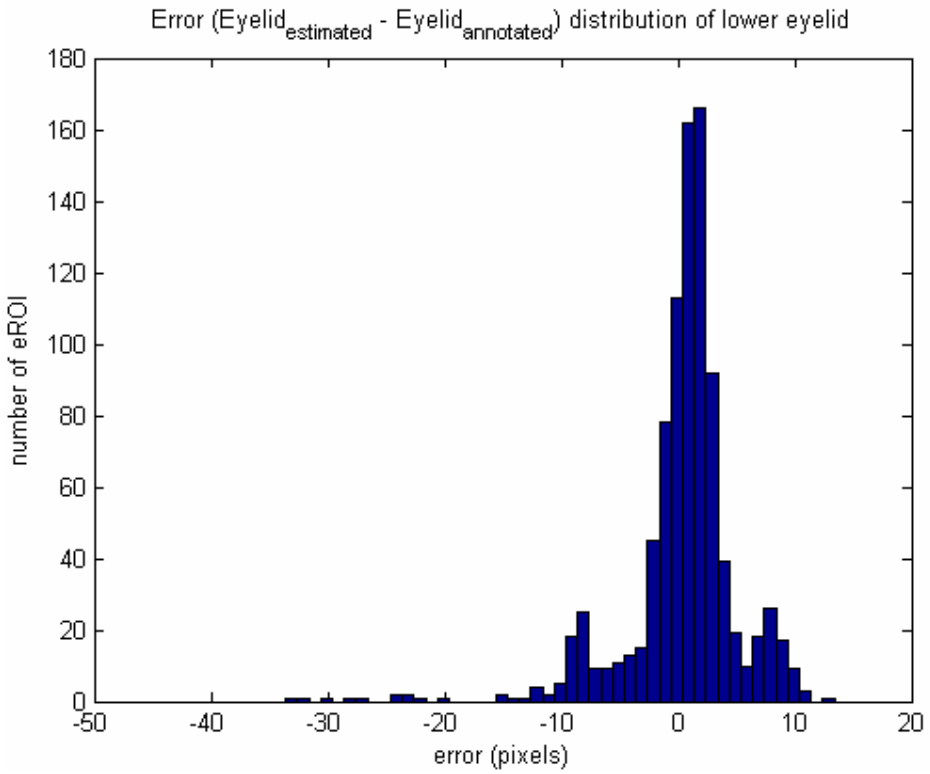


Figure 5-43. Error distribution of the estimated lower eyelid position.

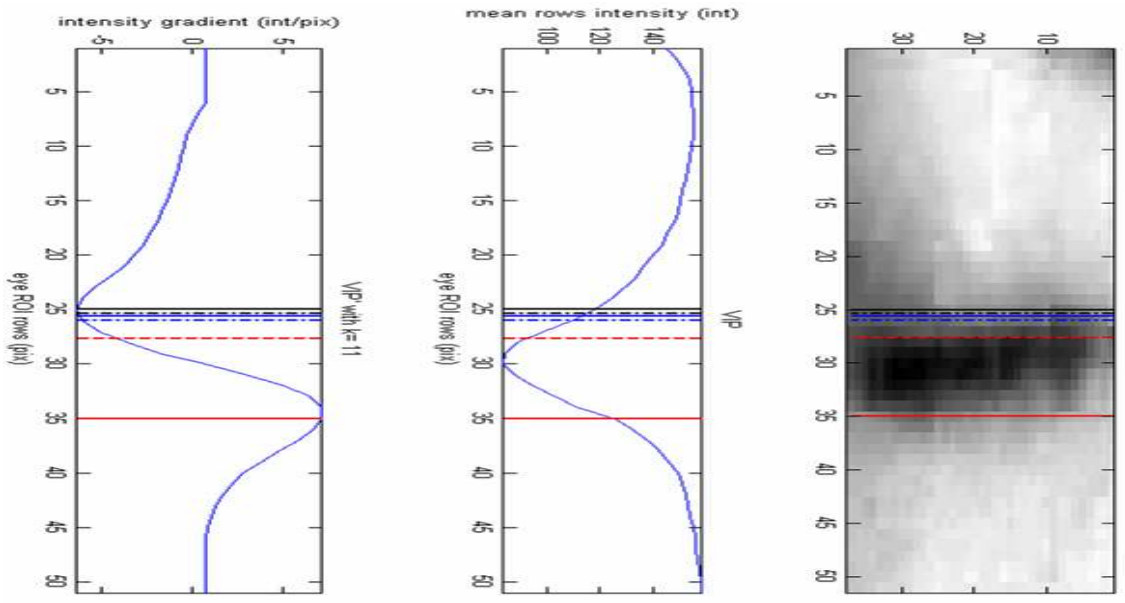


Figure 5-44. An elROI with a fully closed eye in which the row positioned at the bottom of the eyelashes was falsely detected as the LEL_y position.

On the other hand, analysis of the eye images with the LEL_y estimation error less than -5 pixels showed that false detection of the LEL_y above true LEL_y were mainly due to contrast in mean

intensity between features within the iris and contrast in mean intensity between the iris and sclera when the subjects were looking upward, as discussed in section 5.7.3.

The error magnitude in the estimated LEL_y was also evaluated. As shown in Figure 5-45, the lowest mean of 90th percentile error magnitude of 6.3 pixels for the estimated LEL_y was achieved when the width of elROI was set to 69 pixels. On the other hand the lowest mean of median error magnitude of 1.8 pixels for the estimated LEL_y was achieved when the width of elROI was set to 93 pixels. This lowest mean 90th percentile error magnitude (6.3 pixels) of the estimated LEL_y was poor compared to the estimated UEL_y (2.7 pixels) (see section 5.7.5). Therefore, the optimum width of the elROI was set to 37 pixels based on the performance of UEL_y detection rather than the LEL_y detection.

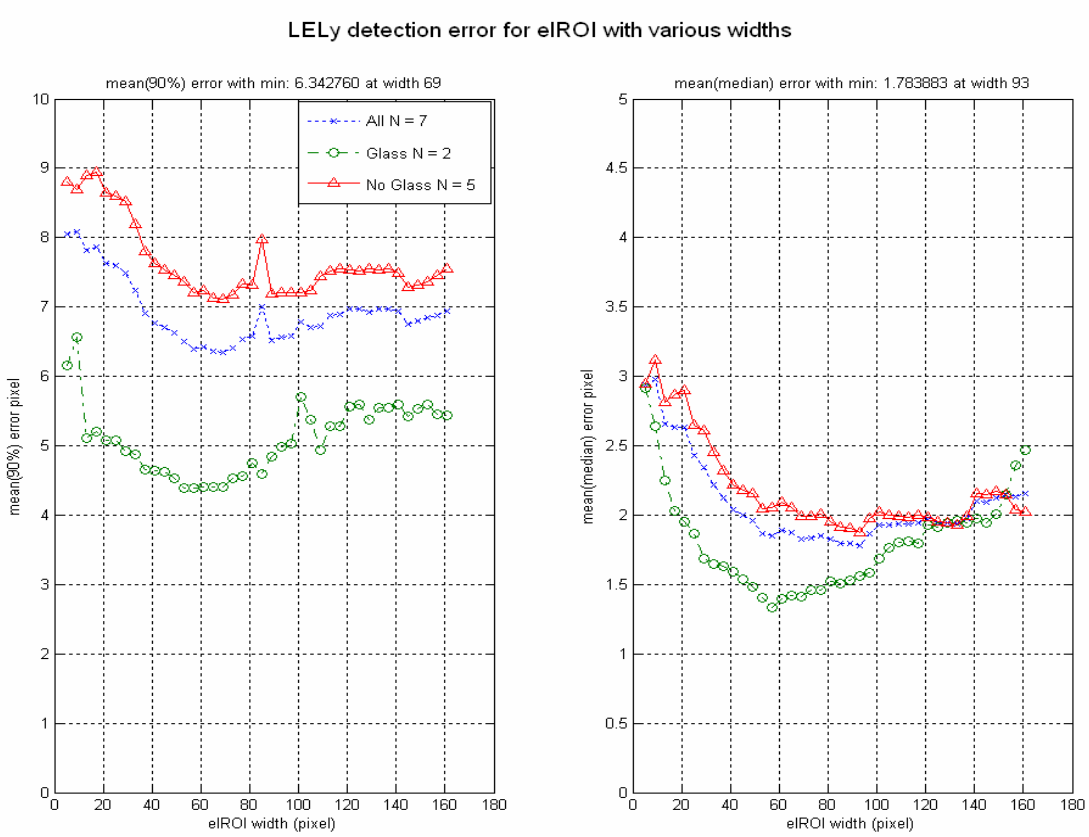


Figure 5-45. Means of 90th percentile and median error magnitude of LEL_y estimation under various widths of elROI.

Table 5-7 lists the means and SDs of 90th percentile error magnitude and median error magnitude of the estimated LEL_y in all 7 subjects when the width of elROI was set to 37 pixels. Here, the mean 90th percentile error magnitude of estimated LEL_y was calculated to be

6.9 pixels, which suggests that, in a worst-case scenario, eye closures can only be measured within 54% of the true measure of eye closure, even if we assume a perfect estimation of UEL_y. The average height between the eyelids of fully open eyes in the annotated reference database was calculated to be 14.2 pixels. However, based on the mean of median error magnitude of LEL_y performance (see Table 5-7), the eye closure can be measured within 86% of the true measure in general if we assume a perfect UEL_y estimation. In other words, general performance of LEL_y estimation is comparable to that of the UEL_y estimation and the worst-case performance of LEL_y estimation is substantially poorer than that of the UEL_y estimation. In conclusion, based on a reliable worst-case performance of LEL_y estimation (90^{th} percentile error magnitude), it was not desirable to use the estimated LEL_y position for measuring eye closure.

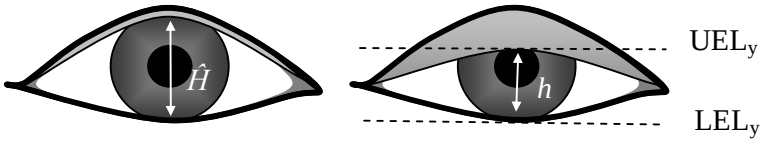
Table 5-7. The means and SDs of the 90^{th} percentile and median error magnitudes of LEL_y estimation in the seven subjects with the width of elROI set to 37 pixels.

Group	Number of subjects (N)	mean ± SD of median error magnitude in LEL_y (pixels)	mean ± SD of 90^{th} percentile error magnitude in LEL_y (pixels)
Subjects with glasses	2	1.6 ± 2.1	4.6 ± 6.8
Subjects without glasses	5	2.3 ± 2.9	7.8 ± 9.8
All subjects	7	2.1 ± 2.9	6.9 ± 2.5

Based on the performance of the upper and lower eyelids detection, it was concluded that only the performance of the UEL_y detection was accurate enough to be used for measuring eye closure. The performance of the LEL_y detection method was inaccurate and unreliable for measuring eye closure.

5.8 Eye-closure measurement

Fractional eye closure (EC) is a measure of the closed portion of an eye relative to when the eye is fully open. Figure 5-46 illustrates the calculation of EC, in which, h is the height of open portion of an eye and $\hat{H}$ is the height when the eye fully open. As eyes close, EC approaches 1.0. To measure EC of an eye in video data, $\hat{H}$ of the eye must first be established and used with the measured h of the eyes in the video.



$$\mathrm{EC} = \frac{\hat{H} - h}{\hat{H}}$$

Figure 5-46. Figure illustrating the calculation of EC. *h* is the height of visible portion of an eye and $\hat{H}$ is the reference height of visible portion of the eye when it is fully open.

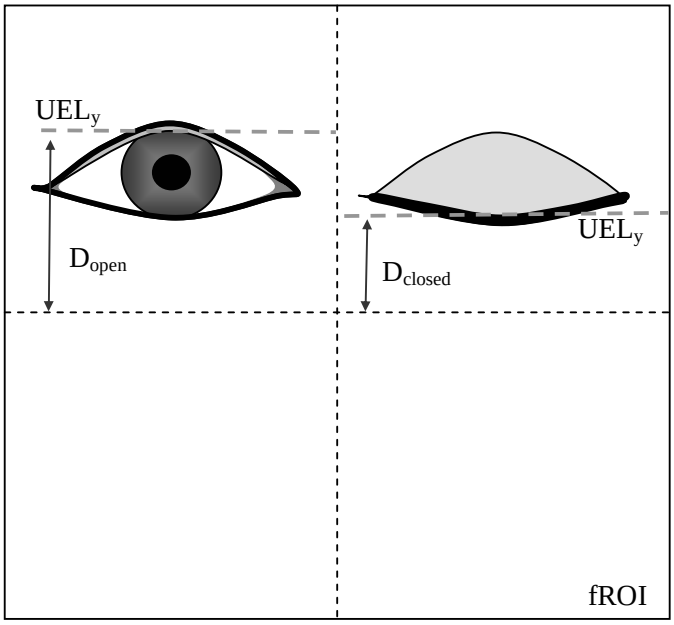
5.8.1 Reference height of an eye

Two methods were investigated for calculating $\hat{H}$ of an eye in a video. In both of these methods, $\hat{H}$ is calculated as a value proportional to the height of the fROI to allow it to be scalable relative to the size of the face in the video. To distinguish these two methods, the $\hat{H}$ values derived from the first and the second methods are labelled as $\hat{H}_a$ and $\hat{H}_b$ respectively.

In the first method, $\hat{H}_a$ was derived by calculating the mean *h* of the eye in a set of frames in which the eye was known to be fully open. *h* of a fully open eye was estimated by taking the difference between the detected UEL$_y$ and LEL$_y$ (section 5.6.5). Although the worst-case performance of the LEL$_y$ detection method was relatively poor (see the mean of 90th percentile error magnitude in Table 5-7), the detected LEL$_y$ was still used for calculating $\hat{H}_a$ because of its reasonable performance in mean of medians error magnitude. In addition, the detection of LEL$_y$ was also observed to be reasonably accurate for most subjects when their eyes were fully open.

To calculate $\hat{H}_a$, a set of 90 calibration frames (equivalent to 3 s of video recordings) with fully open eyes were manually selected for each reference video (section 4.1). The set of calibration frames were collected independently from the set of annotated frames (section 4.2). So, the set of calibration frames for a reference video can include the same frames used in the set of corresponding annotated frames. In the final video-based alertness monitoring system, the frames with fully open eyes can be acquired by initiating a calibration process in which the subjects are asked to keep their eyes fully open for a brief period.

Since the worst-case performance of LEL_y detection was poor, another method was trialled for calculating $\hat{H}_b$. It uses the mean distance between the UEL_y and the y-coordinate of centre of fROI ($fROI_c(y)$) in frames with open eyes and closed eyes, as illustrated in Figure 5-47.



$$\hat{H}_b = D_{open} - D_{closed}$$

Figure 5-47. Schematic illustrating the calculation of $\hat{H}_b$ based on difference between mean D_{open} and D_{closed}, which are calculated from two sets of frames in which the eyes were fully open and fully closed, respectively.

In Figure 5-47, D_{open} is the mean vertical distance between UEL_y and $fROI_c(y)$ in a set of frames in which the eyes are fully open and D_{closed} is a mean vertical distances between the same points in another set of frames in which the eyes are fully closed. Hence, an additional set of 90 calibration frames with fully closed eyes was visually acquired for each of the collected reference video to determine D_{closed}. Theoretically, the method for deriving $\hat{H}_b$ should be better than that for deriving $\hat{H}_a$ because it avoids the use of LEL_y with its poor worst-case performance.

5.8.2 Comparison of methods for reference height calculation

The performance of the two methods for determining $\hat{H}$ was determined by analysing their error magnitudes. To calculate the error magnitudes, the actual mean height of fully open eyes ($\hat{H}_a$) for each set of annotated frames was acquired by calculating the mean distance between

the manually annotated UEL_y and LEL_y in the three annotated frames with fully open eyes in the set. For all 7 subjects, two reference videos (under ambient and dark lighting conditions) were annotated and in each reference video, a set of 33 frames, including the 3 frames with fully open eyes, were annotated (see section 4.2.1).

Figure 5-48 shows the mean error magnitudes of $\hat{H}_a$ and $\hat{H}_b$ for each of the 7 subjects in the reference database. For each subject, the mean error magnitude was derived by calculating the mean of error magnitudes in the left and the right eyes in the two corresponding reference videos of the subject. The mean of the error magnitudes from all 7 subjects was 1.7 pixels for $\hat{H}_a$ and 4.1 pixels for $\hat{H}_b$.

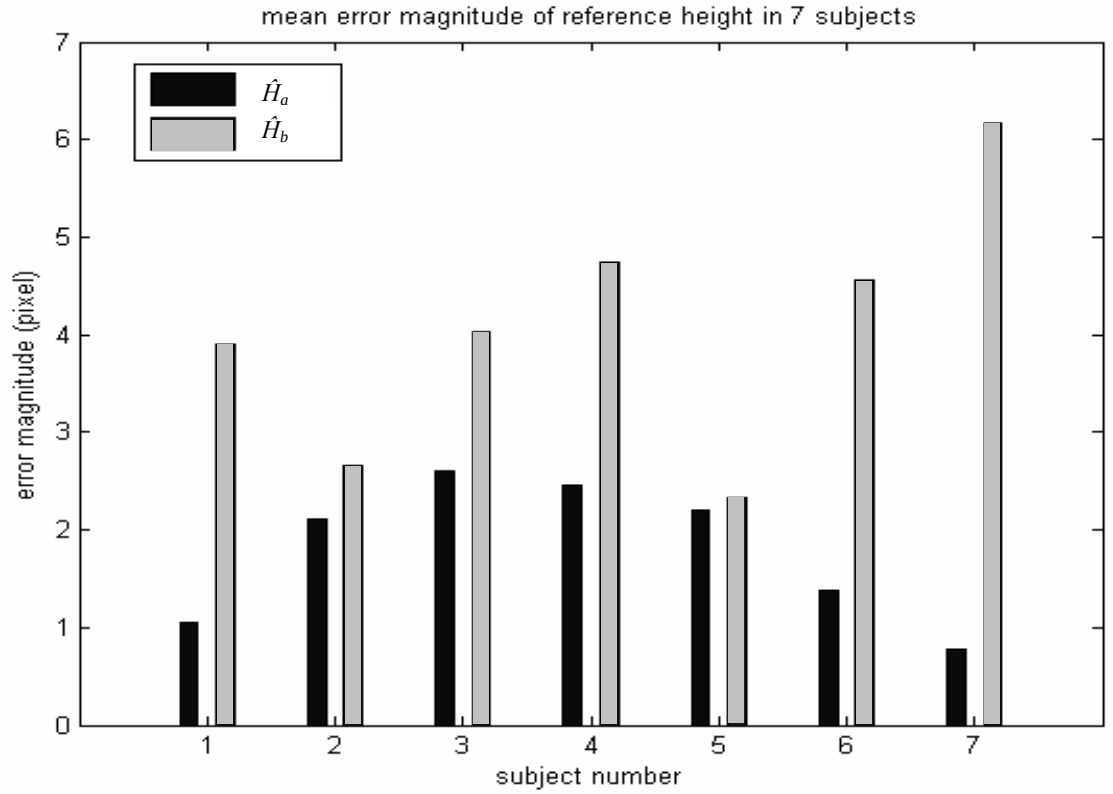


Figure 5-48. Mean error magnitude of $\hat{H}_a$ and $\hat{H}_b$ for the seven subjects in the reference database.

From the evaluation of error magnitudes in Figure 5-48 and its mean, it was concluded that overall the method for estimating $\hat{H}_a$ is better than the method for estimating $\hat{H}_b$. Hence, the method for calculating $\hat{H}_a$ was chosen for estimating $\hat{H}$ to measure EC. The mean error magnitude of 1.7 pixels is a good result for estimating $\hat{H}$, particularly when the mean 90th

percentile error magnitude of the estimated UEL_y (Table 5-6) and LEL_y (Table 5-7) were much higher.

Further analysis showed that the poor performance of $\hat{H}_b$ estimation was mainly due to large errors in detection of UEL_y in calibration frames with closed eyes. Figure 5-49 shows an example of a closed eye calibration frame in which the UEL_y (marked by the upper edge of the ROI with solid line) in both eyes were estimated above their true UEL_y position.

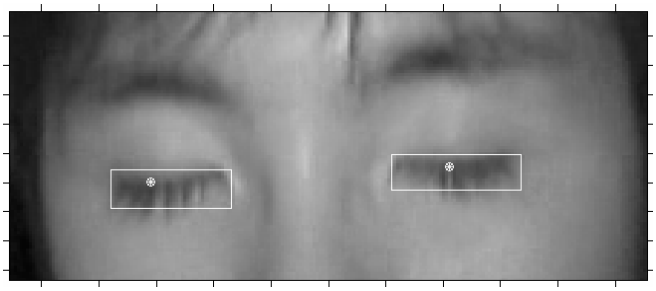
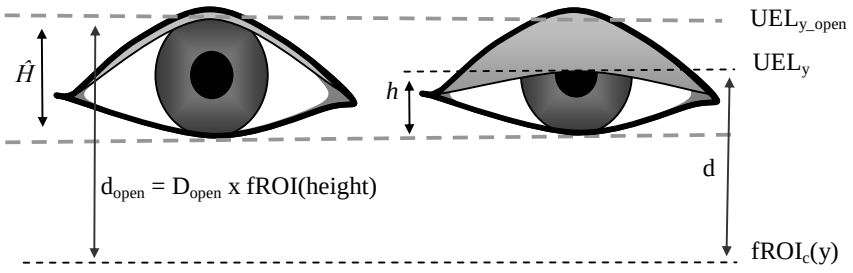


Figure 5-49. An example showing inaccurate estimation of UEL_y in frames with closed eyes. The upper and lower edge of the ROI with solid line represents the UEL_y and LEL_y, respectively. Interestingly, the y-coordinate of the COE marked by '*' appears to be a better indicator of the true UEL_y when the eyes are closed.

However it was interesting to observe, as in Figure 5-49, the y-coordinate of the estimated COE (COE_y) was a better indicator of the true UEL_y in the majority of frames with closed eyes. Using the COE_y instead of the UEL_y in the calibration frames with closed eyes to calculate $\hat{H}_b$ reduced its overall mean error magnitude from 4.1 pixels to 2.5 pixels. The reduction in mean error magnitude of $\hat{H}_b$ further emphasized that the error in $\hat{H}_b$ estimation was mainly due to inaccurate detection of UEL_y.

5.8.3 Fractional eye closure measurement

The fractional eye closure (EC) of an eye in each frame can be measured by calculating the ratio of the height of closed portion of the eye ($\hat{H}$ - h) to the $\hat{H}$ of the eye (see Figure 5-46). The h of an eye is measured based on the estimated UEL_y relative to the mean UEL_y measured when the eye is fully open (UEL_{y_open}), as illustrated in Figure 5-50.

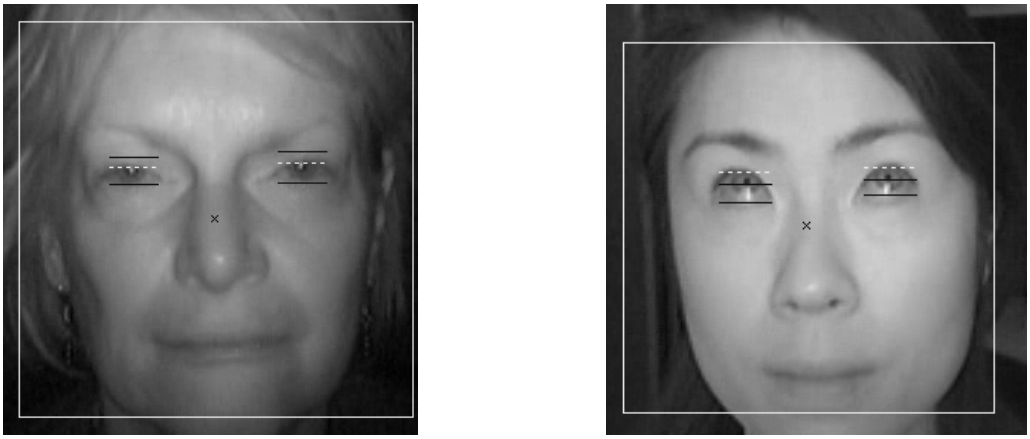


$$\hat{H} - h = \left(d_{open} - d\right) = \left(UEL_{y\_open} - UEL_y\right)$$

Figure 5-50. Illustration of the method for calculating height of the closed portion of an eye ($\hat{H}$ - h) in a frame using the estimated ULE$_y$ relative to the UEL$_y$ when the eye is fully open (UEL$_{y_open}$).

In Figure 5-50, d represents the distance between the estimated UEL$_y$ and fROI$_c$(y). Similarly, d_{open} represents the distance between UEL$_{y_open}$ and fROI$_c$(y). For a given frame, d_{open} is determined by scaling the pre-defined mean proportional distance (D_{open}) relative to the height of the detected fROI. As described in section 5.8.1, D_{open} of an eye for a video is determined by calculating the mean distance between the estimated UEL$_y$ and fROI$_c$(y) in a set of calibration frames with fully open eyes. To make d_{open} consistent for each frame in a video, D_{open} is calculated as a value proportional to the height of the fROI. Once D_{open} of an eye is determined, the height of closed portion of the eye ($\hat{H}$ - h) in each frame is calculated by taking the difference between d and d_{open}. Subsequently, EC is measured by dividing the height of closed portion of the eye by $\hat{H}$.

The value of EC measurement normally ranges from 0 to 1, representing the range of fully open to fully closed eyes, respectively. The fully open eye refers to when the eye is relaxed and naturally open without squinting or forced super-wide open. For example, in Figure 5-51(a) the average EC of both eyes was measured to be 0.38. However, the value of EC can also be negative when the eyes are forced wide open. For example, in Figure 5-51(b) when the subject looks upwards, the eyes open wider than when they are fully open and result in a negative EC value.



(a) (b)

Figure 5-51. Example of eye closure measurement. The two solid lines represent the position (UEL_{y_open}) and reference height ($\hat{H}$) of the eye when it is fully open. The dash line represents the estimated UEL_y. The average EC of the left and right eyes in these frames were (a) EC = 0.38 (b) EC = -0.63.

5.9 Summary

In this chapter, the methods and their performance for detecting the fROI, eROI, COE, and eyelid positions in an image were presented. The EC of an eye in video data was measured by applying these facial feature detection methods. In summary, to measure EC of an eye, a set of frames with fully open eyes was collected from the video and the facial feature detection methods were applied to these frames to determine $\hat{H}$ and D_{open} (section 5.8.1) of the eye. Then, the EC of the eye in every subsequent frame of the video was measured by using the estimated fROI and UEL_y in conjunction with the $\hat{H}$ and D_{open} of the eye (section 5.8.3).

Chapter 6 System performance

This chapter presents a quantitative evaluation and analysis of the eye closure measurement system. Initially, the performance of the overall system is presented and then the system is further analysed in terms of various operational requirements of the video-based alertness monitoring system outlined in section 3.1.

6.1 Reference data for evaluation of eye closure measurements

The performance of the eye closure measurement system was evaluated by analysing its error in estimating fractional eye closure (EC) of the eyes in the reference frames (see section 4.2.1). The error in estimating EC was calculated by subtracting the annotated EC from the estimated EC.

The EC in reference frames was annotated by taking the ratio of the annotated height of the closed portion of the eye to the mean annotated height of an fully open eye ($\hat{H}_{ant}$) (see section 5.8.2). In the reference frames, the heights of closed portion of eyes were derived by subtracting the annotated height of the open portion of the eye from corresponding $\hat{H}_{ant}$. The open portion of the eye was the distance between the annotated UEL_y and LEL_y. Since UEL_y and LEL_y overlap each other when the eyes are fully closed, the open portion of the eye was set to 0 pixels in reference frames with fully closed eyes. Therefore, ECs in reference frames with fully closed eyes were annotated to be 1.

In terms of pixels, the mean $\hat{H}_{ant}$ of the 7 subjects in the reference database was 14.2 pixels. Therefore, if the EC of an eye is measured to be 0.1 or 10% closed, the height of the closed portion of the eye would be approximately 1.4 pixels.

6.2 Requirement of eye closure measurement

For the EC measurement system to be effective for monitoring alertness, it must be able to differentiate the EC in a frame into at least 5 levels of eye closure: closed, ¾ closed, ½ closed, ¼ closed, and fully open. These levels of eye closures correspond to the frame selection used during the reference data collection process in section 4.2.1. To differentiate the eyes into 5

levels of eye closure, the EC of the eyes must be estimated to within an accuracy of ±0.125, as illustrated in Figure 6-1.

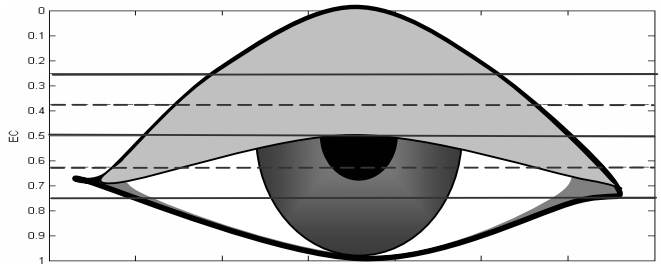


Figure 6-1. Image demonstrating the EC measurement system must determine the measured EC within ±0.125 of the true EC for the system to effectively differentiate eyes into one of the 5 levels of eye closure.

6.3 Eye-closure performance

Figure 6-2 shows the error distribution of the estimated EC of the 924 eyes in 462 reference frames. The EC error is normally distributed with a mean of -0.02 and SD of 0.39.

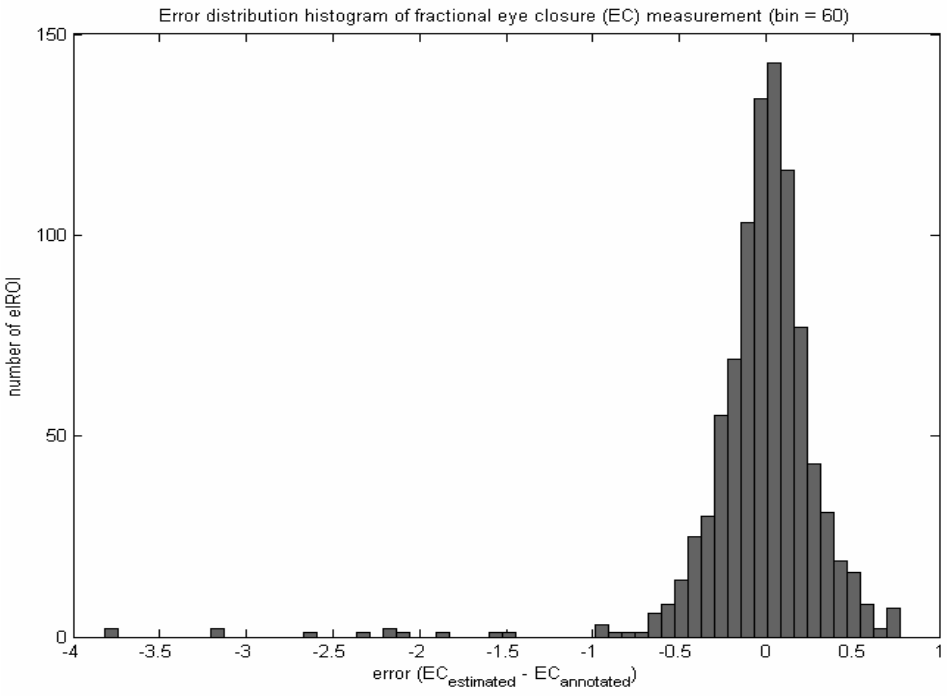


Figure 6-2. Error distribution of the estimated EC in 924 elROI (including both the left and the right eye) in 66 annotated frames (33 under ambient lighting and 33 under dark room) of the 7 subjects in the reference database.

Twelve of the 924 elROI analysed had estimated EC errors of less than -1. The large EC error in these elROIs was due to false detection of eyebrow as the COE, as for the right eye shown in Figure 6-3. Therefore, the accuracy in measurement of the EC of an eye is critically dependent on correct estimation of its COE (section 5.5) and UEL_y. If eyes with EC error of less than -1 are discarded, the mean and SD of the EC error distribution improve to 0.01 and 0.25, respectively.

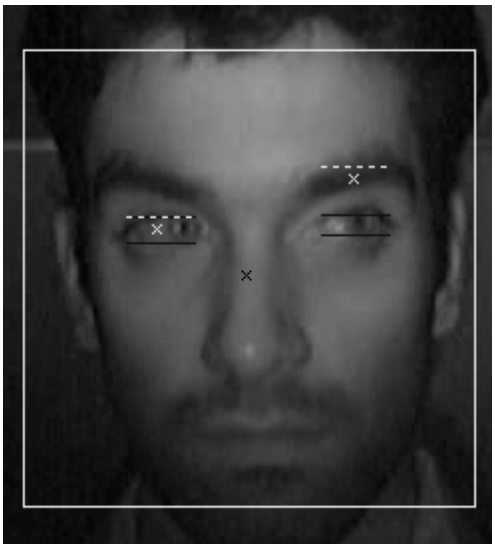


Figure 6-3. Image showing false detection of eyebrows as the COE in the left eye of subject 1. This false detection results in inaccurate detection of UEL_y and subsequently a large error in EC for that eye.

Figure 6-4 shows the median error magnitude and 90th percentile error magnitude for the EC measurement in the reference frames of the 7 subjects. The means and SDs of the error magnitudes of EC are listed in Table 6-1. The general performance (mean of the medians) of the EC measurement system was within 20% of the required accuracy (±0.125) for differentiating the eyes into one of 5 levels of eye closure (see section 6.2). However, the worst-case performance (mean of 90th percentiles) of the EC measurement system is significantly poor and therefore the method is too unreliable for use in monitoring behavioural facial signs of alertness at this stage.

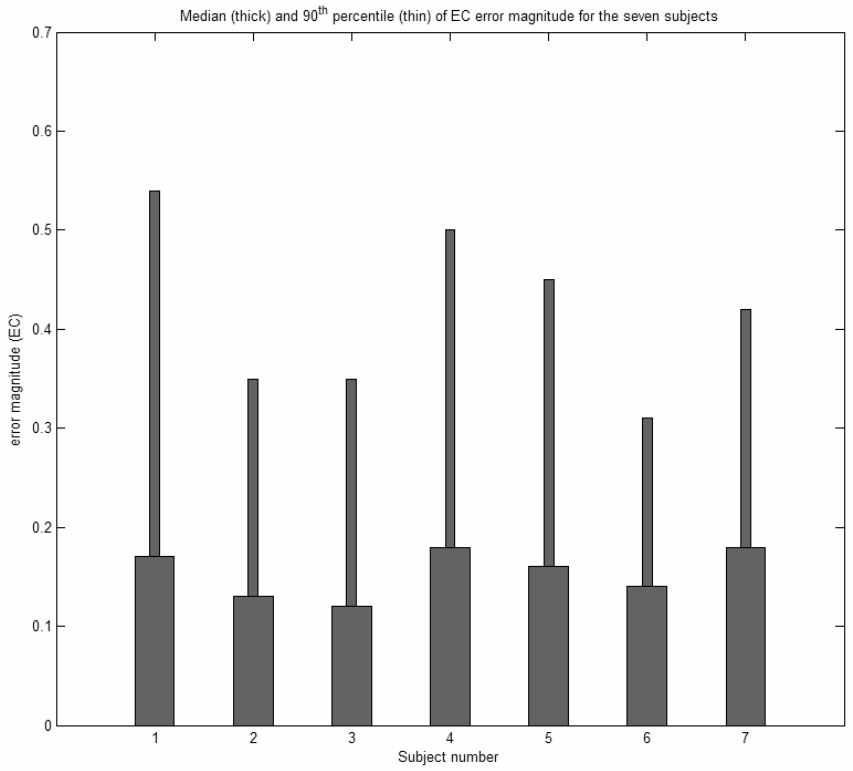


Figure 6-4. The median error magnitude and 90th percentile error magnitude of the EC for the 7 subjects in reference database.

6.3.1 Inter-subject variability

The video-based alertness monitoring system is required to cover a wide range of variations in facial features across the general population (section 3.1). In Table 6-1, the performance of the EC measurement system is presented in terms of factors which contribute to variations in facial features (section 4.1.1) between subjects.

The similarity in the error magnitude of ECs for each subgroup of subjects in Table 6-1 and individual subjects in Figure 6-4 show that the performance of the developed EC measurement system is largely independent of variation in facial features between subjects. This is a good result as it indicates that the EC measurement system is relatively robust to variation in facial features between subjects. However, it should be noted that only 2 of the 4 subjects who wore glasses were used for quantitative evaluation of the facial feature detection methods presented in this report. As explained in section 5.6.5, the 2 subjects who wore glasses (subjects 8 and 9) were rejected because of the difficulty in detecting their eye position due to the artifacts produced by the glasses. Hence, the EC measurement system is less effective in presence of artifacts due to glasses. However, the performance of the EC measurement system was similar

regardless of glasses if the artifacts, such as reflections and dark thick dominant features around the eyes due to certain type of frames, are removed.

6.3.2 Effect of lighting conditions

For the EC measurement system to be practical for monitoring alertness, it must operate effectively under both ambient and dark lighting conditions (section 3.1). Figure 6-5 shows the median error magnitude and 90th percentile error magnitude of the EC measurement under these two lighting conditions. Their corresponding means and SDs are listed in Table 6-1. The general performance of the EC measurement system was 6% more accurate with respect to mean of medians for video recorded in ambient lighting than in a dark room. This improvement in performance is due to higher contrast between the eye features in ambient lighting condition. However, there wasn't any notable variation in the worst-case performance (mean of 90th percentile) of the EC measurement system under the two lighting conditions.

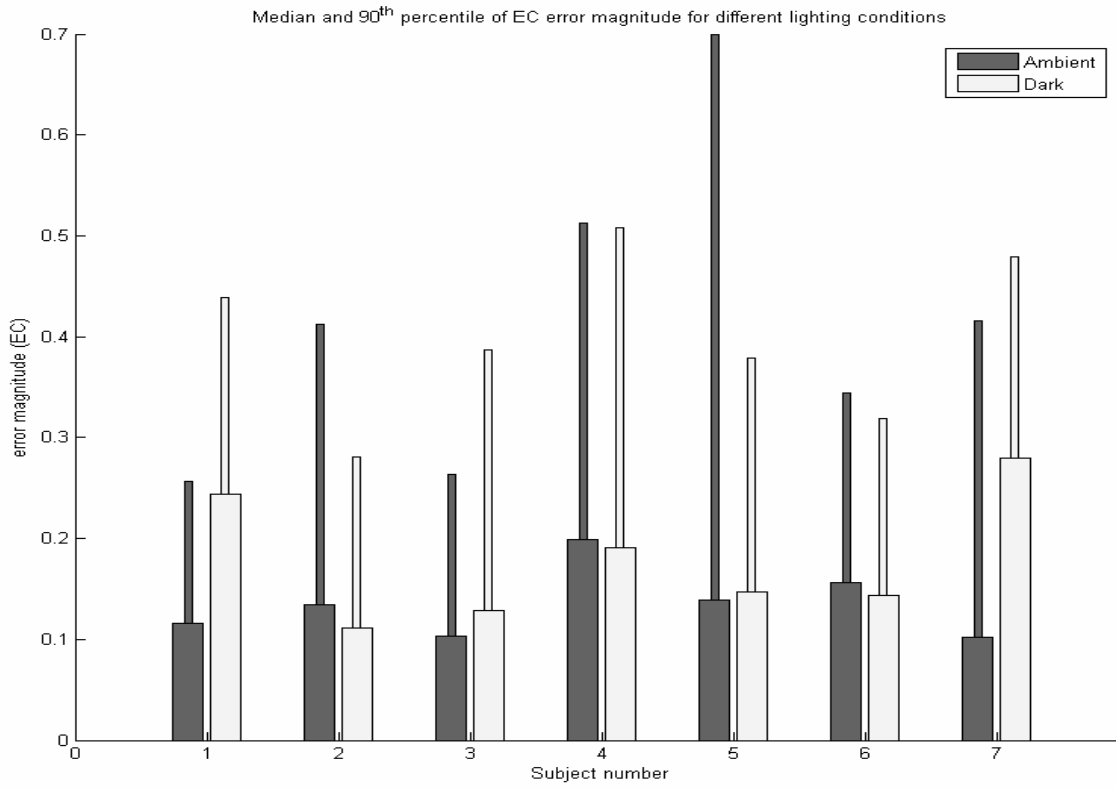


Figure 6-5. Performance of the EC measurement of the 7 subjects for the reference frames recorded under dark and ambient lighting conditions.

Table 6-1. Performance of the EC measurement system in terms of factors that can influence its operation.

Groups	Subject Number	Mean ± SD of median EC error magnitude	Mean ± SD of 90th percentile EC error magnitude
Overall	1 to 7	0.15 ± 0.03	0.42 ± 0.08
Glasses			
Glasses	5, 7	0.17 ± 0.02	0.43 ± 0.03
No Glasses	1, 2, 3, 4, 6	0.15 ± 0.03	0.41 ± 0.10
Gender			
Male	1, 3, 4	0.16 ± 0.05	0.39 ± 0.11
Female	2, 5, 6, 7	0.15 ± 0.05	0.41 ± 0.12
Eyelid folds			
Single	4, 5, 6	0.15 ± 0.02	0.45 ± 0.13
Double	1, 2, 3, 7	0.15 ± 0.06	0.37 ± 0.08
Race			
Occidental (European)	2, 7	0.16 ± 0.07	0.40 ± 0.09
Oriental (Chinese)	4, 5, 6	0.15 ± 0.02	0.45 ± 0.13
Oriental (Indian)	1, 3	0.15 ± 0.06	0.33 ± 0.08
Lighting conditions with NIR illumination			
Ambient light	1 to 7	0.13 ± 0.02	0.41 ± 0.14
Dark room	1 to 7	0.18 ± 0.06	0.40 ± 0.08

6.3.3 Differentiating degree of eye closure

The performance was analysed with respect to 11 categories of eye conditions to determine its ability to correctly differentiate eyes into one of 5 levels of eye closure. The 11 categories of eye conditions included various levels of eye closure and gaze direction as explained in section 4.1.2. Figure 6-6 shows the mean error and the mean error magnitude of EC measurement for the 11 eye conditions. For evaluation of the EC measurement system, the frames with different degrees of eye closures (categories 1-5) are of primary interest. However, the frames with different gaze directions (categories 6-11) also provided information on the performance of EC measurement system. An example of the EC measurement system's performance for a subject (subject 2 in a dark room) under the 11 different eye conditions is presented in Figure 6-7.

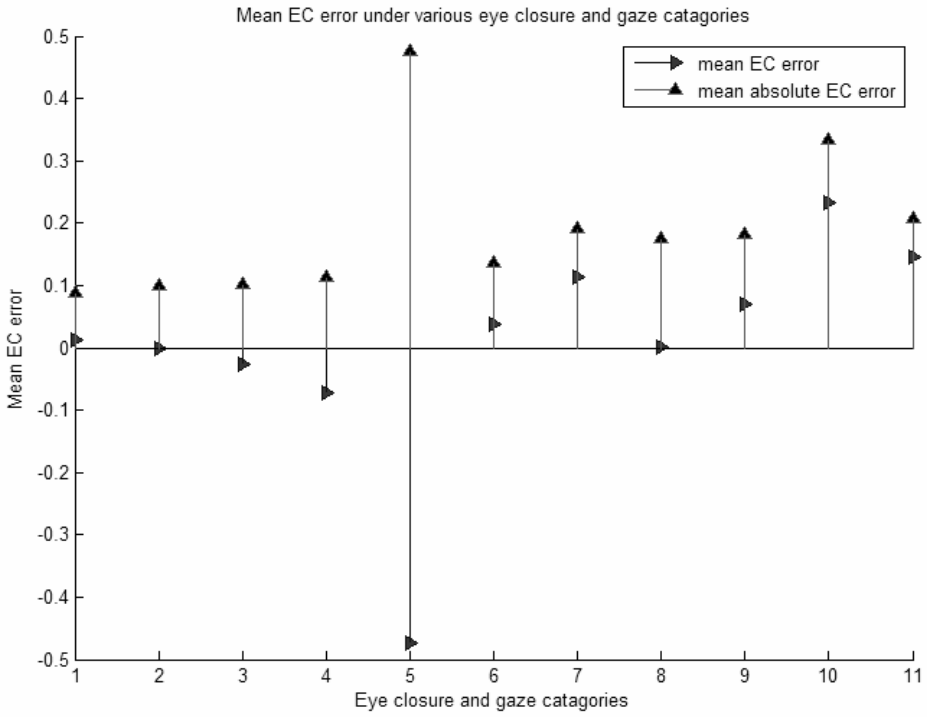


Figure 6-6. Performance of EC measurement method for various eye closures and gaze directions. The category 1 to 11 sequentially represent the frames where the eyes are fully open, ¼ closed, ½ closed, ¾ closed, fully closed, and have left gaze, extended-left gaze, right gaze, extended-right gaze, upward gaze, and downward gaze. Apart from frames with closed eyes (category 5), the mean EC error was relatively small in frames with other 4 degrees of eye closures (category 1-4). In most of the gaze direction categories (category 6-11), the mean EC error were positively biased, i.e., the eyes in these frames were estimated to be more closed than they were. The mean error magnitude (mean absolute EC error) shows that the EC measurement was worst in frames where the eyes were closed, followed by frames where the subjects were looking upwards.

Figure 6-7. Example of EC measurement in annotated frames from subject 2 corresponding to 11 categories of eye condition (see section 4.2.1). The images from the top left to the bottom right represent the eye condition categories from 1 to 11 respectively. The EC is measured based on the position of the UEL_y (dashed line) relative to the reference height ($\hat{H}$) and position (UEL_{y_open}) represented by the two solid lines. The mean EC of the left and right eyes for each of these images were 0.03, 0.14, 0.38, 0.57, 0.98, 0.08, 0.03, 0.04, -0.20, -0.19, 0.83, respectively.

Figure 6-6 shows the mean EC error for the first 4 levels of eye closure (categories 1-4) was within ±0.125. Therefore, as explained in section 6.2, the EC measurement system can differentiate eyes into one of these 4 levels of eye closure when the subject is looking straight ahead.

However, Figure 6-6 also shows that the mean error of EC measurement was considerably high and negatively biased when the eyes were fully closed (category 5). The poor performance of EC measurement system in majority of frames with fully closed eyes is mainly due to incorrect estimation of the UEL_y above its true position as explained in section 5.8.2. In the majority of frames with closed eyes, UEL_y is usually detected at approximately half way between the $\hat{H}$ of the eyes, as shown in Figure 6-8, which results in the eyes estimated as half rather than fully closed. Consequently, it is difficult for the EC measurement system to differentiate fully closed eyes in its current state. Nevertheless, if the temporal information about degree of eye closure in preceding frames is known, it should be possible to identify when eyes are fully closed in video data. On average, the EC measurement system performed better in frames with frontal gaze (categories 1-5) then in frames with other gazes (categories 6-11). This is promising as subjects are most likely to be looking straight ahead when they are drowsy or about to have microsleep.

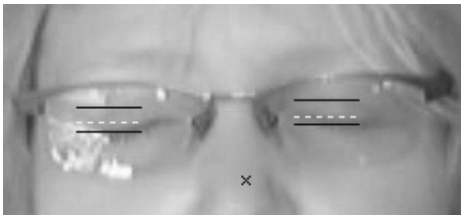
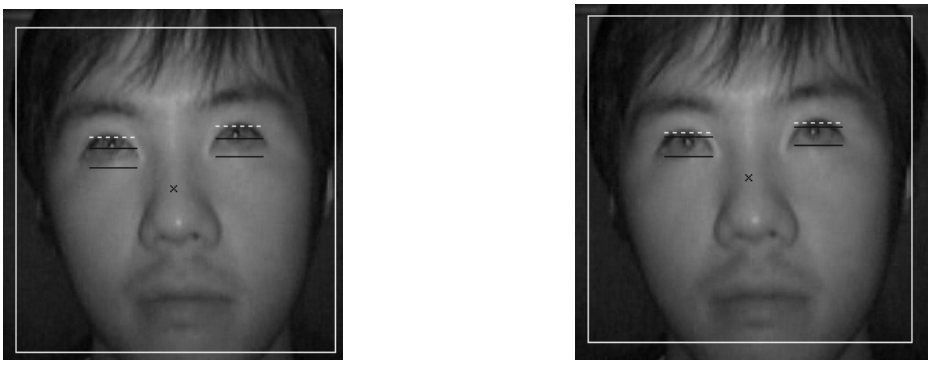


Figure 6-8. An image of closed eyes showing incorrect detection of UEL_y, resulting in the EC being incorrectly measured as 0.68.

Analysis of frames with various gaze directions (categories 6-11) showed that EC measurement was particularly poor and positively biased in frames with upward gaze (category 10). This was mainly due to incorrect estimation of UEL_{y_open}. During upward gaze, the position of the palpebrae fissure of subjects also moves upward relative to the face. However, the detected fROI does not move. Since, the UEL_{y_open} is relative to the position and height of the fROI (see section 5.8.3), the UEL_{y_open} is incorrectly estimated and consequently the EC is inaccurately measured. Figure 6-9(a) shows as example of incorrect estimation of UEL_{y_open}, which results

in poor measurement of EC. However, in some frames, the fROI also readjusts with the upward gaze, as shown in Figure 6-9(b), resulting in more accurate measurement of the EC.



(a) EC = -0.53 (b) EC = - 0.15

Figure 6-9. Measurement of EC is dependent on the accuracy in estimation of UEL_{y_open}. (a) The eyes in this frame have large EC measurement error due to poor estimation of UEL_{y_open}. (b) In this frame, the error in EC measurement is reduced because fROI moved upward during upward gaze resulting in correct estimation of the UEL_{y_open}.

In addition to frames with upward gaze, any displacement of fROI produced large error in EC. For example, although the UEL_y and $\hat{H}$ are correctly estimated in Figure 6-10, displacement of fROI results in incorrect estimation of the UEL_{y_open} and consequently a large error in EC.

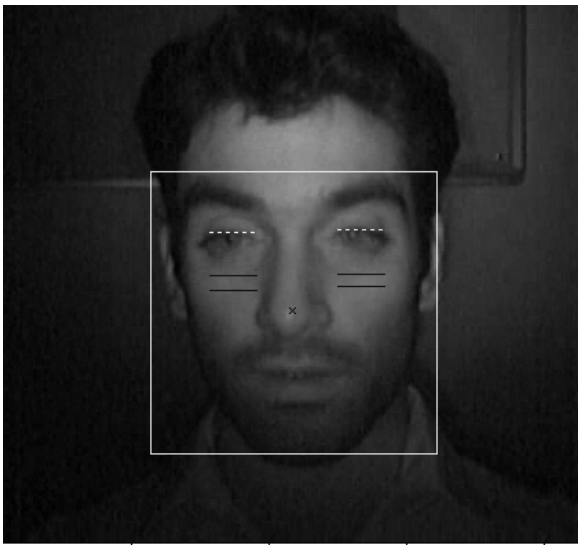


Figure 6-10. The detected fROI in this frame was lower and smaller than its neighbouring frames. This displacement of the fROI resulted in incorrect estimation of the UEL_{y_open} and a large error in EC.

6.4 Sources of error in eye closure measurement

The error in EC measurement is a result of accumulated errors in estimation of UEL_y, $\hat{H}_{open}$, and UEL_{y_open}. Based on the sum of error magnitudes in estimation of the UEL_y and $\hat{H}$ (see sections 5.7.5 and 5.8.2), an error in EC of 3.1 pixels on average and 4.4 pixels in worst-case scenario would be expected. Since the average $\hat{H}_{ant}$ in the reference database was 14.2 pixels (refer to section 6.1), the means of the median error magnitude and the 90th percentile error magnitude of the EC would be expected to be at least 0.21 and 0.31, respectively. As these expected errors for EC are already relatively high without even accounting for the error in estimation of the UEL_{y_open}, it was not surprising to see the poor worst-case performance (average 90th percentile of error magnitude = 0.42) of the EC measurement system. The error in estimation of the UEL_{y_open} could not be quantified because no stationary marker relative to the face was annotated.

The two main sources of larger errors during UEL_y estimation were the false detection of the eyebrow as the COE (see Figure 6-3) and the incorrect detection of UEL_y in frames with closed eyes (see Figure 6-8). The error in estimation of the $\hat{H}$ was mainly due to poor localization of the LEL_y (refer to section 5.7.3) in the calibration frames with fully open eyes. Figure 6-11 shows an example of a poorly estimated $\hat{H}$ of the left eye, which eventually results in inaccurate measurement of EC for the eye. As explained in section 6.3.3, the error in estimation of the UEL_{y_open} is mainly due to the displacement of fROI.

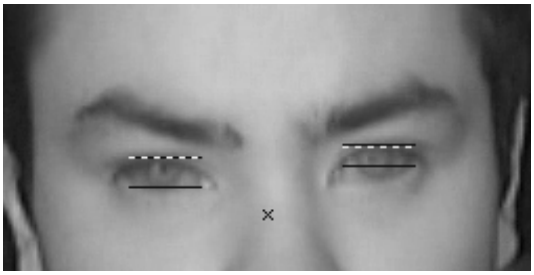


Figure 6-11. In this image the $\hat{H}$ for the left eye is under estimated, resulting in the left eye being estimated to be more closed than it is.

Errors in the EC could be considerably reduced by improving the localization of the LEL_y. One method for improving LEL_y is suggested later in section 7.3.1.1. With better localization of LEL_y, $\hat{H}$ can be estimated more accurately and also the UEL_{y_open} can be eliminated for measurement of EC. Reduction in error from these two parameters will improve the overall performance and reliability of the EC measurement system.

6.5 General remarks

Although the individual average performance of the facial feature detection methods used in this project were encouraging, the compound errors of these methods resulted in unsatisfactory performance of EC measurement for monitoring alertness. This is one of the main disadvantages of the passive top-down system development approach adopted in this project.

The EC measurement system in its current state has a general accuracy (average median error magnitude = 0.15) sufficient to differentiate the eyes into 3 levels of eye closures, i.e., fully open, half open, and fully closed. However, its worst-case performance (average 90^{th} percentile error magnitude = 0.42) is too unreliable for differentiation to any particular level of eye closure. Therefore, the EC measurement system in its current form cannot be reliably used for automatically detecting the facial behavioural signs of drowsiness and microsleeps. However, these results were promising for this initial proof of concept attempt. With improvement in localization of the LEL_y the performance of the EC measurement system could be considerably improved.

Chapter 7 Discussion, conclusions, and future work

This project was initiated with the aim of developing a real-time video-based system for automatically measuring facial metrics to identify visible facial signs of drowsiness and behavioural microsleeps. Three significant facial metrics were identified – eye closure, eye movement, and head movement (Chapter 1). Measurement of these facial metrics required the development of algorithms to automatically detect the positions of the face and the eye features including the apex of eyelids, centre of pupil/iris, and eye corners. In addition, for the developed facial metrics measurement system to be practical and user compliant, it had to be non-intrusive, insensitive to lighting conditions, and tolerant to variation in facial features (Chapter 3).

In this final chapter, the main contributions of the thesis towards achieving the initial objectives of the project are summarized. Where possible, the performances of the methods developed are compared to that of similar existing systems. Limitations of the methods developed and suggestions for further improvements and future developments are discussed.

7.1 Summary of main contributions and findings

7.1.1 Remote camera-based non-intrusive system

As described in Chapter 3, a remote camera-based system with an NIR-illumination source and NIR-pass filter was developed using low-cost off-the-shelf hardware for monitoring alertness. Unlike head-mounted or head-gear systems, a remote camera-based system allows non-intrusive monitoring of a subject in their natural operating environment without movement constraints. Use of the NIR-illumination source and filter makes the system insensitive to visible lighting conditions and allows it to operate in both well-lit and completely dark conditions without introducing visual distractions to subjects. A critical requirement was that the alertness monitoring system be able to operate under very low levels of visible illumination because losses of responsiveness are most likely to occur under these conditions.

However, unlike a head-mounted system, a remote camera-based system requires additional computer-vision algorithms to localize the face and eyes within an image. These algorithms add computational load and inaccuracies to the system. In addition, the distance between the

camera and the subject considerably reduces the resolution and quality of the eye-image, which, in turn, reduces the precision of the eye features detection methods and consequently the measurement of facial metrics. In hindsight, it would have been better to have used a higher quality camera, specifically designed for machine-vision applications and with a much higher resolution and optical quality than a webcam for greater precision in measurement of facial metrics. However, the reasonable performance achieved with the webcam in this project was encouraging to support the purchases of a higher quality camera for future development.

7.1.2 Reference image database

To quantitatively evaluate the developed methods, reference videos of nine subjects were recorded under four different lighting conditions (Chapter 4). These videos contain various gaze directions, eyelid movements, and head movements, representing the behavioural signs of drowsiness and microsleeps. From two video recordings of every subject a set of frames was selected and manually annotated for multiple eye and facial features. In addition to allowing the determination of image characteristics and system performance in this project, the reference database will be a valuable resource for evaluation of methods developed in future. To our knowledge there are no publicly available reference image databases recorded under NIR illumination and specifically containing images exhibiting signs of drowsiness.

7.1.3 Development of eye-closure measurement system

In this project, algorithms for detecting positions of the face, eyes, and eyelids were developed and quantitatively evaluated. Of the three desired facial metrics (eye-closure, eye movement, and head movement), the facial-feature algorithms are only able to measure eye closure at this stage. However, the algorithms can contribute to future work aimed at measuring eye movement and face movement.

Measurement of eye closure allows the detection of blinks and prolonged eye closure, which are the prominent facial signs of drowsiness and microsleeps (Chapter 1 and 3). The eye-closure measurement system was developed using a top-down passive feature-detection approach which involved processes that are summarized as follows:

- using a Kalman filter to stabilize and track the fROI output acquired from an existing Haar-face detection algorithm,

- deriving proportional anthropometric constants for localization of eROI relative to the fROI,
- forming an eye template and detecting the centre of eye position using the eye-template matching method,
- developing an eyelids detection algorithm based on the vertical integral projection of the eROI, and
- deriving a method for measuring eye closure to achieve 3 levels of eye closures (fully open, half closed and fully closed).

Dependency of a method's performance on preceding methods proved both a blessing and a curse in the top-down passive feature-detection approach adopted in this project. The success of the complete eye-closure algorithm depends on each serial part functioning well. For example, improvement in the stability of the fROI detection with a simplified Kalman filter improved the localization of the eROI and consequently the detection of eye features. Likewise, improvement in detection of the COE with Gaussian correction matrix based on prior knowledge considerably improved the performance of the eyelid detection algorithm.

Conversely, compounding of errors in UEL_y, $\hat{H}$, and UEL_{y_open} components of the eye-closure measurement system resulted in poor 90^{th} percentile accuracy in measures of eye closure. However, the median accuracy of the eye-closure measurement system was within 20% of the required accuracy and was sufficient to distinguish between fully open, half closed, and fully closed eyes. The system was also found to be tolerant to variations in facial features between subjects, except for certain artifacts due to spectacles.

One of the main sources of error in measurement of eye closure was poor reliability of the LEL_y at the 90^{th} percentile accuracy. Hence, future work on improving the accuracy of LEL_y detection should improve the performance of the eye-closure measurement system. Although, the eye-closure measurement system requires further development to be suitable for alertness monitoring, the relatively high performance of the UEL_y detection algorithm provided an encouraging result.

7.2 Comparison with other systems

The parameters used for reporting the performance of relevant computer vision algorithms (see Chapter 2) are inconsistent and, hence, it is difficult to make direct comparisons with the methods developed in this project. There are also substantial differences in the characteristics of images used to determine the performance of various computer vision methods. Therefore, the facial feature detection methods developed in this project, which were using an in-house reference image database, cannot be directly compared with other systems/methods in the literature. Nonetheless, an attempt has been made to compare performance with other methods.

The AntiSleep system (Smart Eye AB, Sweden) (section 2.5) can measure the distance between eyelids with a 2 mm accuracy, together with head position, head orientation, and gaze direction. The average error magnitude in eyelids detection in this project was 3.5 pixels. As a pixel in the reference database is approximately 1.3 mm, the average accuracy of the distance between the eyelids in this thesis equates to 4.5 mm. The error in eyelid distance comes predominantly from poor detection of lower eyelid. The median accuracy of the upper eyelid detection (1.4 pixels) equates to 1.82 mm and 90^{th} percentile accuracy (2.4 pixels) equates to 3.51 mm. This suggested that improved detection of the lower eyelid has the potential to achieve a similar level of eye closure to that of the AntiSleep system.

Another remote camera-based fatigue and drowsiness detection system that is in an advanced stage of research has been developed by Ji et al. (Ji & Bebis, 1999; Ji & Yang, 2001, 2002; Ji et al., 2004; Zhu & Ji, 2005). They determine the eye position by detecting the retinal reflection of the NIR illumination that produces a bright pupil and measure eye closure based on the changing elliptic shape of the pupil. In their system, the RMS error in detection of the centre of pupil is reported as 1.09 and 0.68 pixels for x and y coordinates, respectively. The RMS error in COE detection in our system was calculated to be much higher at 4.2 and 2.1 pixels for the respective x and y coordinates. However, the precision in detection of eye position in our system is not as critical as in system developed by Ji et al. because in our system the COE was primarily detected so as to differentiate the y-coordinate of the eye for detecting eyelid positions.

Ji et al. have also reported an average RMS error of 0.08 in measurement of the ratio of radii of ellipses that fit the detected pupil. This is effectively their version of eye closure measurement because as eyelids close and occlude the pupil, the shape of the ellipse fitting the pupil also changes. In comparison, the average error in fractional eye closure (EC) in our system was

calculated to be 0.21. However, it should be noted that the size of the pupil varies as the iris dilates or constricts in response to the intensity of the visible light. Hence, measurement of eye closure based on the occlusion of the pupil by the eyelids is not directly comparable to the measurement of the actual aperture of the eyelids measured in this thesis. In addition, a study on a driver-vigilance monitoring system, based upon NIR retinal reflection (Bergasa et al., 2006), has reported that the performance of bright pupil detection was substantially reduced in daylight conditions when the pupil was smaller and the noise-to-signal ratio high.

However, in future, the combination of both the passive and active NIR-illumination based systems could be considered for detecting eye movement and gaze direction if alternative methods for detecting center of pupil (see section 7.3.3) produce poor results. In addition to the development of facial feature detection methods, Ji et al. has also carried out further research (Ji et al., 2004) into utilizing these video-based facial metrics to derive metrics of level of alertness. Closure scrutiny of their alertness metrics is likely to be beneficial in future developments of a video-based alertness monitoring systems.

Video-based alertness monitoring systems are quickly becoming a strong application of computer-vision algorithms (Bergasa et al., 2006; D'Orazio, Leo, & Distante, 2004; Eriksson & Papanikolopoulos, 2001; Grace, 2001; Ji et al., 2004; Singh & Papanikolopoulos, 1999; Smith et al., 2003; Z. Tian & Qin, 2005; Zhu & Ji, 2005). Particularly, there is a substantial interest in this application from automotive industry (Desai & Haque, 2006; Ji & Bebis, 1999; Ji & Yang, 2002; Lal & Craig, 2001; Ueno et al., 1994). However, the high cost of commercially available products and a need for more robust computer-vision algorithms continue to motivate further research in this area. In addition, the development of robust, reliable, and accurate facial-feature detection algorithms have application far beyond sleep-related studies (see Chapter 2). Hence, further research and development in the field of computer-vision based facial-feature detection algorithms is well justified.

7.3 System limitations and future work suggestions

The current EC measurement system has several limitations and it requires further work if it is to be sufficiently reliable for monitoring alertness. In this section, the limitations and suggestions for methods to overcome those limitations are discussed. Methods for measurement of head and eye movement are also suggested.

7.3.1 Improvements to eye-closure measurement system

7.3.1.1 Improved lower eyelid detection

Currently the EC measurement system can at best only identify 3 levels of eye closure. Ideally, the system would be able to identify at least 5 levels of eye closure. As suggested in section 6.4, the precision of EC measurement could be substantially improved with better accuracy of LEL_y detection in frames with open eyes. More reliable and accurate LEL_y detection would reduce the errors in estimation of $\hat{H}$ and eliminate the errors from ULE_{y_open} estimation by making it unnecessary for eye closure measurement.

False detection of the features within the iris (see section 5.7.3) is the main reason for error in LEL_y detection in frames with open eyes. A simple way to improve LEL_y detection would be to use a wider elROI when detecting the LEL_y. As shown in mean of medians performance in Figure 5-45, the LEL_y detection method has lowest error when the width of the elROI is set to 93 pixels. A wider elROI would integrate the contrast in intensity of larger area of the eye image into the *VIP*, hence making the contrast in intensity within the iris less dominant. Figure 7-1 shows an example of a substantial improvement in estimation of $\hat{H}$ of subject 3 when the LEL_y in the calibration frames was detected with wider elROI (previously the width of the elROI was set to 37 pixels and the new width was set to 93 pixels). Unfortunately, time restrictions prohibited a full quantitative evaluation to assess the improvement in LEL_y detection for the entire database. However, it would be desirable to further investigate the effect of this simple change in the overall performance of the eye closure measurement.

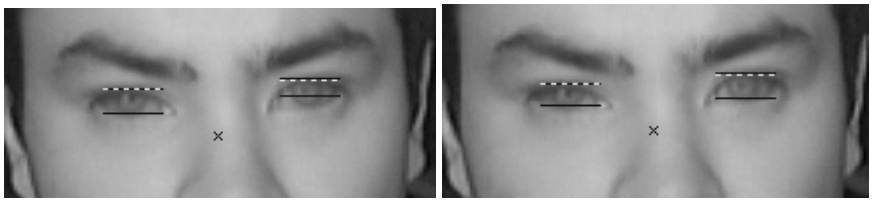


Figure 7-1. The detection of LEL_y can be improved by using wider elROI, which in turn would improve the estimation of $\hat{H}$. The LEL_y in the open eye calibration frames in the above two images were estimated by setting the width of the elROI to Left: 37 pixels and Right: 93 pixels. The estimation of $\hat{H}$ in the left eye is considerably improved.

7.3.1.2 Use of temporal information

Currently, the measurement of EC is unreliable in frames with closed eyes. However, EC measurement is reasonably good when the eyes are fully or partially open. Therefore, the

incorporation of temporal information by way of preceding frames as the eyelids close would allow the frame with the fully closed eyes to be more reliably identified.

7.3.1.3 Dynamic sizing of the eye template

Currently, a fixed-size eye template is used for detecting the COE. However, the size of the eye template could be dynamically scaled in proportion to the size of the fROI and, hence, allow COE detection method to better accommodate a varying facial size in the image as the subject moves back and forth. Nevertheless, the use of a fixed-size eye template for COE detection had minimal effect on overall performance presented in this thesis as subjects were seated at a fixed distance from the camera during the collection of the reference video data.

7.3.1.4 Reducing the effects of spectacles

Two of the four subjects (i.e., subjects 8 and 9) who wore spectacles were rejected for performance evaluation of methods developed in this thesis. UEL_y detection for subject 8 was poor because reflection off his glasses completely saturated and occluded the upper part of his eyes. Reflections off glasses could not be avoided in an uncontrolled outdoor environment. Hence, under these circumstances, no system would be able to detect features completely occluded in the image. On the other hand, subject 9 was rejected because of the thick dark frame of his spectacles being incorrectly detected as COE. This false detection considerably degraded the 90^{th} percentile accuracy for this subject. Moreover, the Gaussian correction matrix (section 5.6.3) was ineffective because the eye and the frame of glasses were at the opposite edge of the eROI. The median accuracy of COE detection for subject 9 was comparable to other subjects in the reference database. Although, there was some reflection off glasses of the spectacles of other two subjects (subjects 5 and 7), most part of their eyes were clearly visible and both spectacles had a thin metallic frames. The amount of NIR that gets reflected off the glasses of spectacles is partly dependent on the property of the glass material. However, the placement of the camera can be controlled to reduce the effect of the reflection.

7.3.1.5 Improved face ROI detection

The Haar-face detection algorithm incorporated in this system can only localize the fROI with small deviation (approximate elevation of $\pm 20^{\circ}$ and lateral deviation of $\pm 10^{\circ}$) from straight ahead orientations (Fasel et al., 2005). In addition, the proportional anthropometric constants for localizing the eROI were derived relative to the frontal face images. Hence, in the current

system, the fROI and eROI can only be reliably localized in frontal face images. Ideally, detection of fROI within substantial deviation in head orientation would make the system more reliable and robust. However, this is probably not critical in a vehicle-based alertness monitoring system as subjects will, for the most part, be facing and looking straight ahead, especially when they are drowsy and likely to have microsleeps.

Research is being undertaken into angular Haar-object classifiers to allow detection of rotated objects (Barczak, 2005; Lienhart & Maydt, 2002). Angular Haar-object classifiers could be integrated into the Haar-face detection algorithm used in the current system to detect the fROI even with the head deviating laterally or vertically from straight ahead orientation.

7.3.1.6 Detection of head nods

In addition to stabilizing the fROI in consecutive frames, the Kalman filter also tracks the fROI, which allows the system to estimate the fROI if the Haar-face detection algorithm fails for a brief period of time. Furthermore, temporal information from tracking the fROI could also be used to identify head nods commonly seen during behavioural microsleeps. As can be seen in Figure 7-2 (reproduced from Figure 5-11 for convenience), during each of the 9 head nods performed by a subject to imitate deep drowsy behaviour, the y-coordinate of the fROI produces a distinctive pattern which could be used to identify head nods.

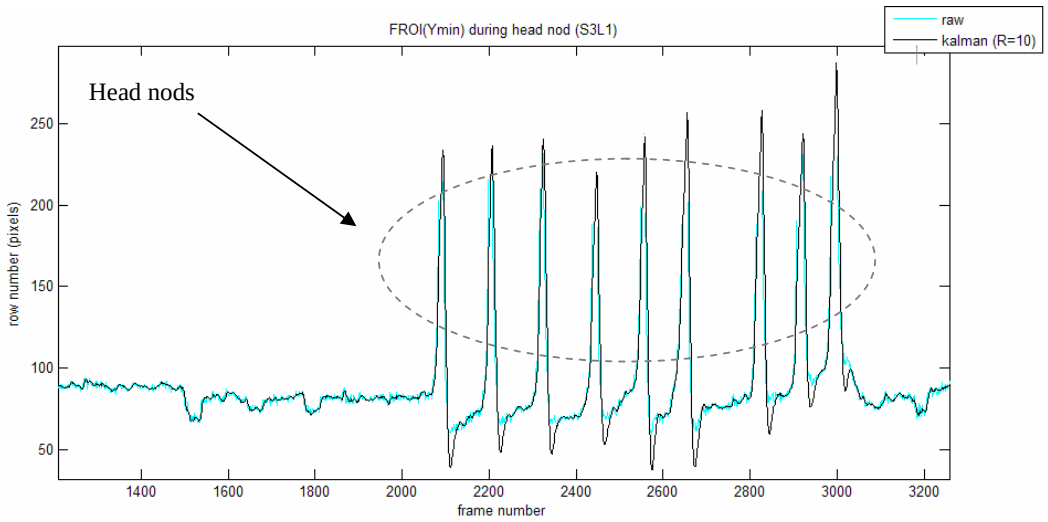


Figure 7-2. During head nod the y-coordinate of the fROI produces characteristic patterns.

7.3.2 Measurement of head movement

The CAM-SHIFT algorithm (Bradski, 1998) which can be readily implemented from OpenCV libraries, could also be integrated into the Haar-face detection so as to obtain up to 4-degrees of freedom (X, Y, Z, and roll) of head motion information. Figure 7-3 shows an example of output of CAM-SHIFT application in OpenCV, in which the angle of the cross on the face corresponds to the side-ways head roll angle.

Figure 7-3. Example of the CAM-SHIFT head orientation output (Bradski, 1998) which can be used to obtain the angular head orientation.

7.3.3 Measurement of eye position and eye gaze

Measurement of horizontal eye movements and determination of gaze direction relative to head position requires detection of the centre of pupil/iris and of the eye corners. Comprehensive development and evaluation of algorithms to detect these eye features were not possible within the scope of this project but iris detection method based on disk eye template is presented in section 3.5.4 and a method for detection of eye corners detection is presented next.

7.3.3.1 Detection of eye corners

Eye corners can be detected by applying the same principles as for eyelid detection in section 5.7. Once the eyelid positions are detected, the elROI can be further narrowed down to optimize eye corner detection. Then, the left and the right eye corners can be detected by localizing the first minimum and last maximum , respectively, in the gradient of the horizontal integral projection of the new eye corner ROI, as shown in Figure 7-4.

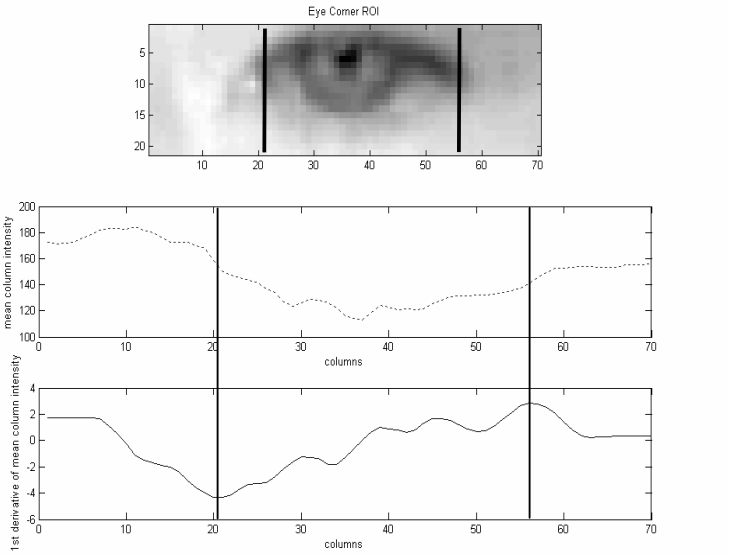


Figure 7-4. Example of detection of eye corners using cues in the horizontal integral projection of the eye corner ROI.

7.3.4 Implementation of a real-time system

An ultimate requirement of the alertness monitoring system, as outlined in Chapter 3, was to be able to operate in real-time. However, development of reliable and accurate measurement of facial metrics was of higher priority than processing speed in this stage of system development. Currently, the eye-closure measurement system does not operate in real-time. All experimental results presented in this thesis were based on post-processing of selected frames. The most computationally intensive part in the current system is the cross-correlation calculation for COE detection (section 5.6.2), which takes approximately 800 ms per frame. Based on this processing speed and trial experience, the current system can operate at approximately 0.5 fps.

In future, code optimisation and integration of the tracking algorithms such as Lucus-Kanade tracking algorithm (Bouguet, 1999) and Kalman filters to track a detected object could be used to increase the processing speed to real-time. In most feature detection systems (Bergasa et al., 2006; Sirohey & Rosenfeld, 2001; Zhu & Ji, 2004b) the real-time operation (25 fps or greater) was achieved by detecting the feature of interest during the initialization process or in frequent intervals and then tracking the feature in the rest of the frames of the video.

7.4 Conclusions

Video-based alertness monitoring systems can be used to carry out further research on the physiological signs of microsleep as well as provide a practical means of monitoring behavioural signs of alertness. The work presented in this thesis forms a firm foundation for future development of a practical, non-intrusive, and light-insensitive video-based alertness monitoring system. The reference video database and the sets of manually annotated frames will also be of value as a resource for further development and evaluation of methods.

The relatively accurate face and eye detection methods used in this project provide foundations for the development of methods to measure the three primary facial metrics of drowsiness and microsleep. The eyelid detection methods developed show encouraging initial results. In particular, detection of the upper eyelid was reasonably accurate and robust. In contrast, poor detection of the lower eyelid substantially reduced the reliability of eye closure measurement. Hence, further investigation into improving the accuracy and reliability in of the lower eyelid detection is strongly recommended to improve the eye closure measurement system to a level suitable for reliably monitoring alertness levels.

Appendix A Software flow diagrams

Flow diagrams of the software developed in each stage of this project (section 3.6) are listed in this appendix. The name of the corresponding source codes written in Matlab and C files are also listed.

A.1 Data acquisition and annotation software

Figure A - 1 shows the flow diagram of the software developed for video data acquisition and annotation of selected frames.

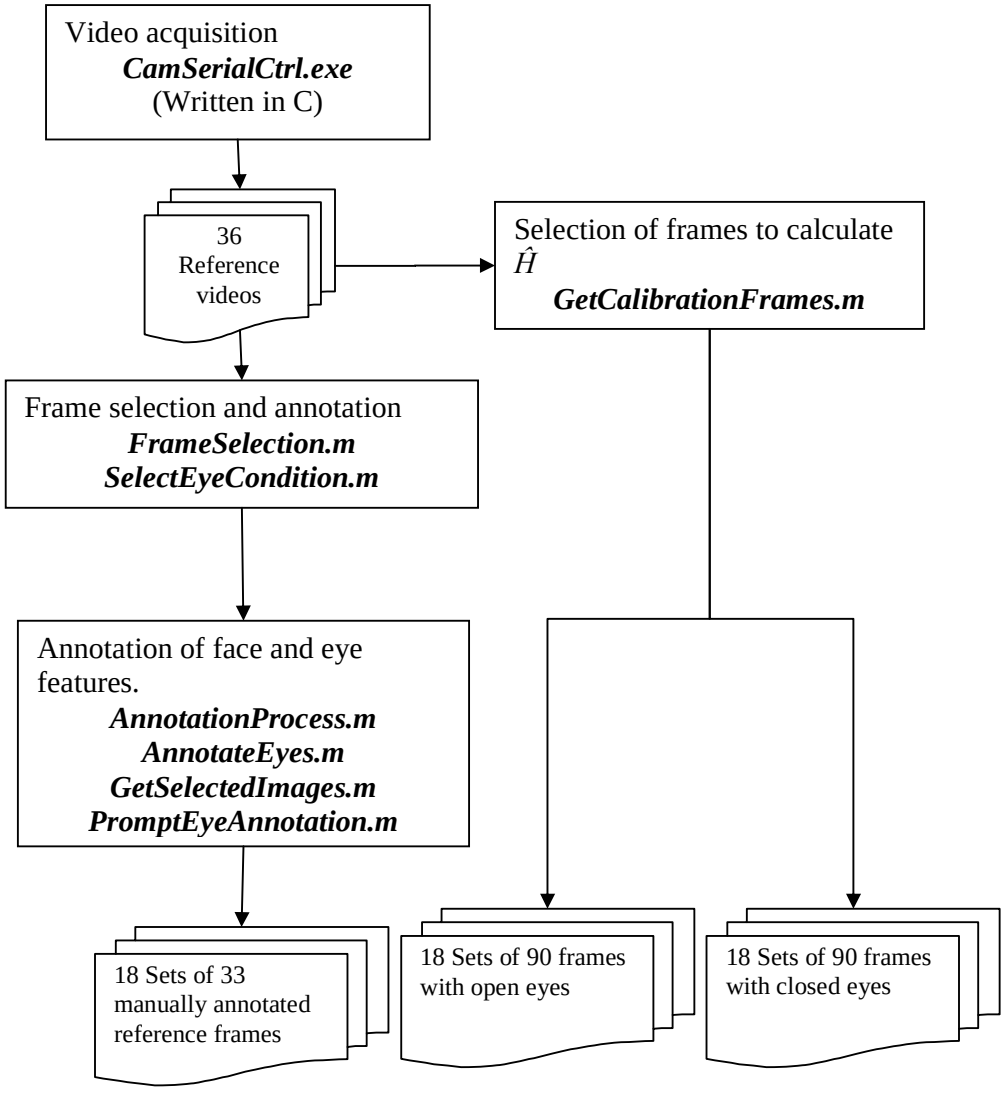


Figure A - 1. Flow diagram of the data acquisition and facial feature annotation software.

A.2 Facial-feature detection software

The selected reference frames were processed offline using the facial feature detection software as shown in Figure A - 2. The set of frames with the known opens eye were also processed for facial detection and then the height between the upper and lower eyelids in these frames were used to determine the $\hat{H}$ (section 5.8.1) of the eyes in the corresponding video.

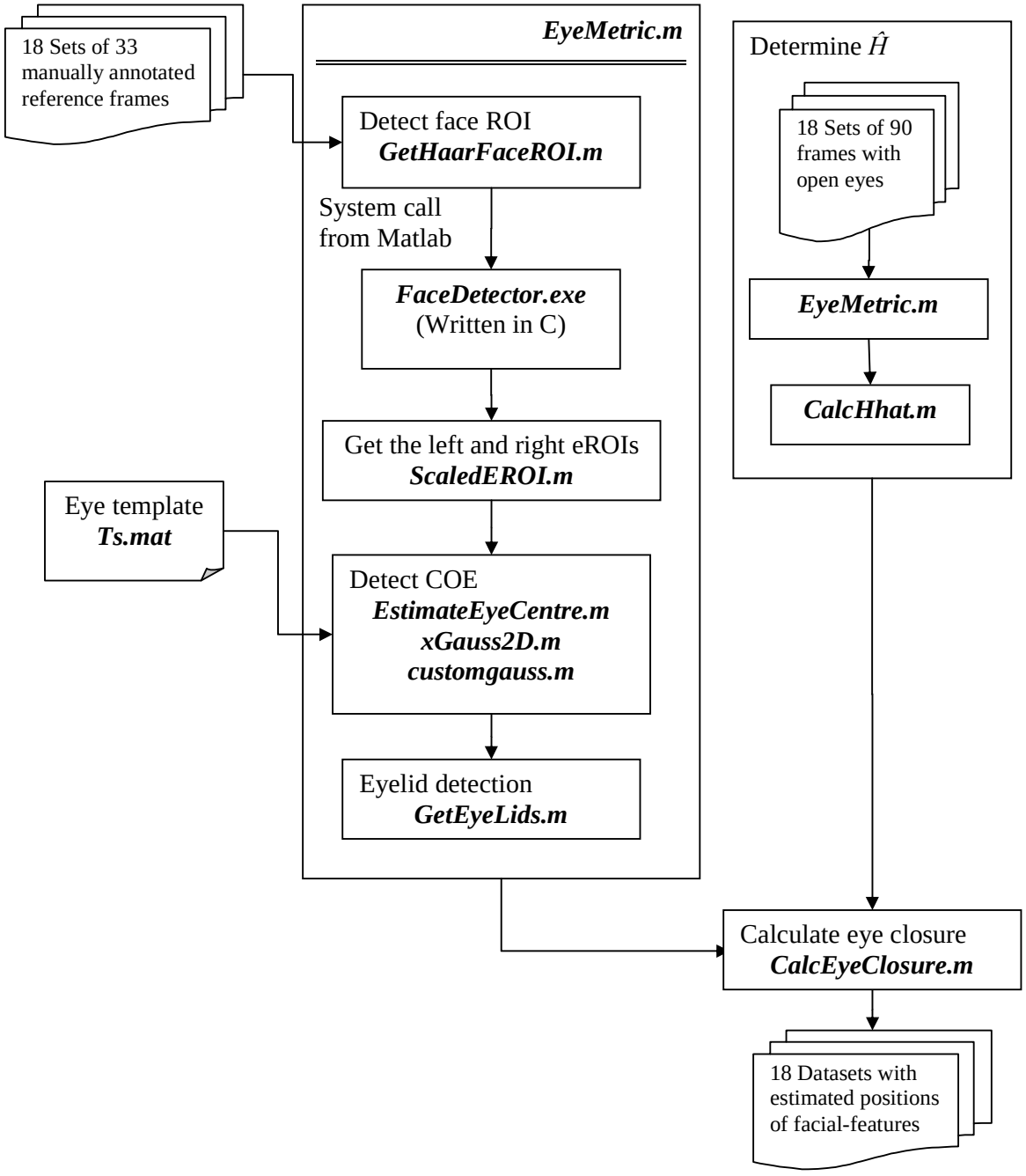


Figure A - 2. Flow diagram of facial feature detection software.

A.3 Evaluation unit software

An evaluation unit was developed to determine the performance of the facial-feature detection algorithms in this project. The annotated data and the estimated data of the facial-features are feed to the evaluation unit in which their error statistics are calculated. Figure A - 3. illustrate this and list the name of the software developed for facial feature performance evaluation.

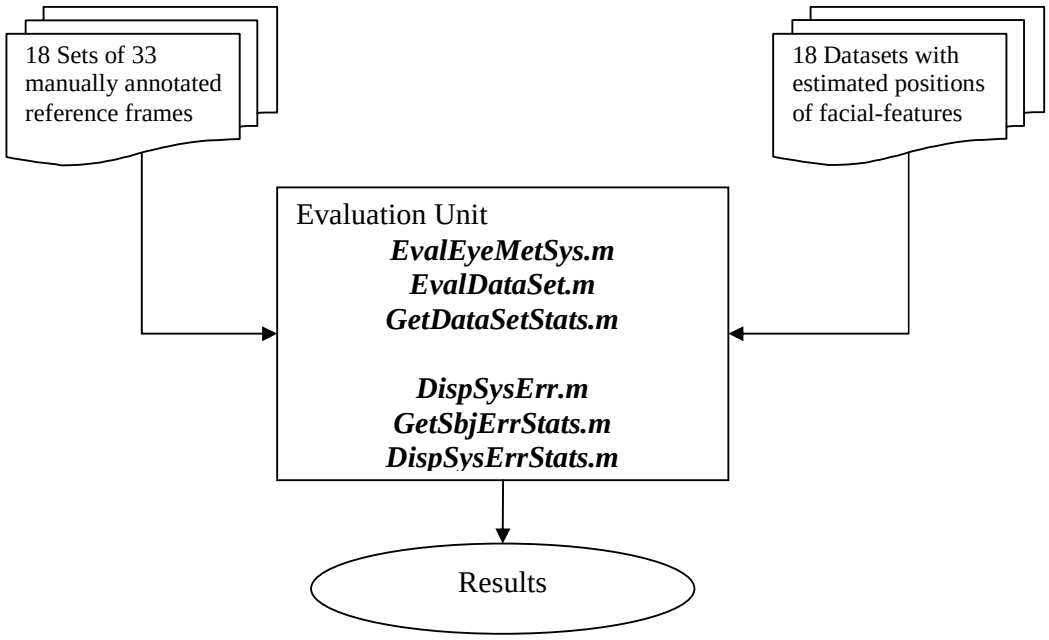


Figure A - 3. List of software for the evaluation unit.

References

AntiSleep. (2005). *Smart Eye AntiSleep - monitoring fatigue and attention cues in automotive applications*. Goteborg: Smart Eye AB, Sweden.

Babcock, J. S., & Pelz, J. B. (2004). *Building a lightweight eyetracking headgear*. Paper presented at the Proceedings of the 2004 symposium on Eye tracking research \& applications, San Antonio, Texas.

Barbato, G., Ficca, G., Beatrice, M., Casiello, M., Muscettola, G., & Rinaldi, F. (1995). Effects of sleep deprivation on spontaneous eye blink rate and alpha EEG power. *Biological Psychiatry, 38*(5), 340-341.

Barczak, A. L. C. (2005, 28-29 November). *Toward an Effecient Implementation of a Rotation Invariant Detector Using Haar-like Features.* Paper presented at the IVCNZ, Dunedin, New Zealand.

Barna, P., & Schlanger, S. (2004). *Fandamentals of the infrared physical layer* (No. AN243): Microchip Technology Inc.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems, 7*(1), 63-77.

BioID. (2001). *The BioID face database*, from http://www.bioid.com/download/facedb/facedatabase.html

Bouguet, J. Y. (1999). Pyramidal Implementation of the Lucas Kanade Feature Tracker

Description of the algorithm. *OpenCV, Microcomputer Research Lab, Santa Clara, CA, Intel Corporation.*

Bradski, G. R. (1998). Computer Vision Face Tracking For Use in a Perceptual User

Interface. *Intel Technology Journal, Q2*(OpenCV, Microcomputer Research Lab, Santa Clara, CA, Intel Corporation).

Brown, R. G., & Hwang, P. Y. C. (1997). *Introduction to Random Signals and Applied Kalman Filtering* (Third ed.): John Wiley & Sons, Inc.

Cohn, J., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). "Automatic recognition of eye blinking in spontaneously occurring behavior". *Behavior Research Methods, Instruments, and Computers, Vol. 35*, 420 - 428.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(1), 21-30.

Davidson, P. R., Jones, R. D., & Peiris, M. T. R. (2005). Detecting behavioral microsleeps using EEG and LSTM recurrent neural networks. *27th Annual International Conference of the Engineering in Medicine and Biology Society*, 5754-5757.

Davidson, P. R., Jones, R. D., & Peiris, M. T. R. (2007). EEG-based lapse detection with high temporal resolution. *IEEE Transactions on Biomedical Engineering in press*.

DeCarlo, A. D., Metaxas, A. D., & Stone, A. M. (1998). *An anthropometric face model using variational techniques.* Paper presented at the Proceedings of the 25th annual conference on Computer graphics and interactive techniques.

Desai, A. V., & Haque, M. A. (2006). Vigilance monitoring for operator safety: A simulation study on highway driving. *Journal of Safety Research, 37*(2), 139-147.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., et al. (1997). Cumulative sleepiness, mood, disturbance and psychomotor vigilance performance decrements during a week of sleep restriced to 4-5 hours per night. *Sleep, 20,* 267-277.

D'Orazio, T., Leo, M., Cicirelli, G., & Distante, A. (2004). *An algorithm for real time eye detection in face images.* Paper presented at the Proceedings of the 17th International Conference on Pattern Recognition.

D'Orazio, T., Leo, M., & Distante, A. (2004). *Eye detection in face images for a driver vigilance system.* Paper presented at the Intelligent Vehicles Symposium, 2004 IEEE.

Duchowski, A. T. (2003). *Eye Tracking Methodology Theory and Practice* (1 ed.). London: Springer-Verlan London Limited.

Ebisawa, Y. (1995). *Unconstrained pupil detection technique using two light sources and the image difference method.* Paper presented at the Visualization and Intelligent Design in Engineering and Architecture Conference.

Ebisawa, Y., & Nurikabe, Y. (2006). *Improvement of PupilMouse.* Paper presented at the Proceedings of the 23rd IEEE Instrumentation and Measurement Technology Conference.

Ebisawa, Y., & Satoh, S.-i. (1993). *Effectiveness of pupil area detection technique using two light sources and image difference method.* Paper presented at the Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE.

El-Bakry, H. M. (2001). *Fast iris detection using neural nets.* Paper presented at the Electrical and Computer Engineering, 2001. Canadian Conference on.

Eriksson, M., & Papanikolopoulos, N. P. (2001). Driver fatigue: a vision-based approach to automatic diagnosis. *Transportation Research Part C: Emerging Technologies, Volume 9*(Issue 6), Pages 399-413.

Evinger, C., Manning, K. A., & Sibony, P. A. (1991). Eyelid Movements, Mechanisms and Normal Data. *Investigative Ophthalmology & Visual Science, 32*(2), 387-400.

Farkas, L. (1994). *Anthropometry of the head and face*: Raven Press.

Fasel, I., Fortenberry, B., & Movellan, J. (2005). A generative framework for real time object detection and classification. *Computer Vision and Image Understanding, 98*(1), 182-210.

Feng, G. C., & Yuen, P. C. (1998). Variance projection function and its application to eye detection for human face recognition. *Pattern Recognition Letters, 19*(9), 899-906.

Forsyth, D., & Ponce, J. (2003). *Computer vision : a modern approach*. London: Prentice Hall.

Foucher, J. R., Otzenberger, H., & Gounot, D. (2004). Where arousal meets attention: a simultaneous fMRI and EEG recording study. *Neuroimage, 22*, 688-697.

Galley, N., & Schleicher, R. (2002). *Fatigue indicators from the electrooculogram - AWAKE consortium internal report.*

Grace, R. (2001). *Drowsy Driver Monitor And Warning System.* Pittsburgh, Pennsylvania: Pittsburgh, Pennsylvania.

Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J.-M., Gricourt, D., Davis, B. K., et al. (1998). *A drowsy driver detection system for heavy vehicles.* Paper presented at the 17th DASC. The AIAA/IEEE/SAE Digital Avionics Systems Conference.

Grandjean, E. (1979). Fatigue in industry. *British Journal of Internal Medicine, 36*, 175-186.

Grauman, K., Betke, M., Gips, J., & Bradski, G. R. (2001). *Communication via eye blinks - detection and duration analysis in real time.* Paper presented at the Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Gu, H., & Ji, Q. (2004). *An automated face reader for fatigue detection.* Paper presented at the 6th IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.

Gu, H., Su, G., & Du, C. (2003, 26-28 November). *Feature Points Extraction from Faces.* Paper presented at the Image and Vision Computing NZ, Palmerston North, New Zealand.

Harrison, Y., & Horne, J. A. (1996). Occurrence of 'microsleeps' during daytime sleep onset in normal subjects. *Electroencephologram Clinical Neurophysiology, 98*, 411-416.

Heitmann, A., Guttkuhn, R., Acacia, A., Udo, T., & Martin, M.-E. (2001). *Technologies for the monitoring and prevention of driver fatigue.* Paper presented at the International driving symposium on human factors in driver assessment, training and vehicle design, Colorado, USA.

Horne, J. A., & Reyner, L. A. (1995). Sleep related vehicle accidents. *British Medical Journal, 310*, 565-567.

Huang, D. Y. (2001). The Drowsy Driver. *Jacksonville Medicine, March.*

Ishikawa, T., Baker, S., Matthews, I., & Kanade, T. (2004). *Passive Driver Gaze Tracking with Active Appearance Models.* Paper presented at the Proceedings of the 11th World Congress on Intelligent Transportation Systems.

Ji, Q., & Bebis, G. (1999). *Visual Cues Extraction for monitoring Driver's Vigilance.* Paper presented at the First HONDA Initiation Grant Forum: The HIG Symposium.

Ji, Q., Wechsler, H., Duchowski, A., & Flickner, M. (2005). Special issue: eye detection and tracking. *Computer Vision and Image Understanding, 98*(1), 1-3.

Ji, Q., & Yang, X. (2001). *Real Time Visual Cues Extraction for Monitoring Driver Vigilance.*Unpublished manuscript.

Ji, Q., & Yang, X. (2002). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging, 8*, 357-377.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology, 53*(4), 1052-1068.

Jin, K.-S., Cho, S.-G., Lee, J.-S., & Hwang, J.-J. (2004). *Real-time pupil detection based on three-step hierarchy.* Paper presented at the Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1991). *Principles of neural science* (4th ed.): McGraw-Hill.

Kapoor, A., & Picard, R. W. (2002). *Real-time, fully automatic upper facial feature tracking.* Paper presented at the Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on.

Karson, C. N. (1992). In the blink of an eye. *Biological Psychiatry, 32*(6), 467-468.

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research, 5*(3), 243-246.

Karson, C. N., Goldberg, T. E., & Leleszi, J. P. (1986). Increased blink rate in adolescent patients with psychosis. *Psychiatry Research, 17*(3), 195-198.

Kawato, S., & Tetsutani, N. (2004). Detection and tracking of eyes for gaze-camera control. *Image and Vision Computing, 22*(12), 1031-1038.

Kolstad, J. L. (1990). *Grounding of the US Tankship Exxon Valdez on Bligh Reef, Prince William Sound Near Valdez, AK, March 24, 1989*. Washington, DC: National Transportation Safety Board, Washington.

Kuo, C. J., Huang, R. S., & Lin, T. G. (1997). *Synthesizing lateral face from frontal facial image using anthropometric estimation.* Paper presented at the IEEE International Conference on Image Processing (ICIP'97), Santa Barbara, CA, USA.

Lal, S. K. L., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology, 55*(3), 173-194.

Lal, S. K. L., & Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology, 39*(3), 313-321.

Lam, K.-M., & Yan, H. (1996). *An Improved Method for Locating and Extracting the Eye in Human Face Images.* Paper presented at the Proceedings of the 13th International Conference on Pattern Recognition.

Lienhart, R., Kuranov, A., & Pisarevsky, V. (2002). *Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection*. Santa Clara, CA 95052, USA: Intel Labs, Intel Corporation.

Lienhart, R., & Maydt, J. (2002). *An extended set of Haar-like features for rapid object detection.* Paper presented at the Image Processing. 2002. Proceedings. 2002 International Conference on.

Liu, X., Xu, F., & Fujimura, K. (2002). *Real-time eye detection and tracking for driver observation under various light conditions.* Paper presented at the IEEE Intelligent Vehicle Symposium.

Makito, S., Mitsuo, S., & Minoru, N. (1997). Blink detection from pupil images of a driver's face : (Mitsubishi Electric Corp.). *JSAE Review, 18*(2), 204.

Mallis, M. M. (1999). Evaluation of Techniques for Drowsiness Detection: Experiment on Performance-Based Validation of Fatigue-tracking Technologies. *Drexel University.*

Marcus, C., & Loughlin, G. (1996). Effect of sleep deprivation on driving safety in house satff. *Sleep, 19*, 763-766.

Matthes, R. (2000). ICNIRP statement on light-emitting diodes (LEDS) and laser diodes: Implications for hazard assessment. *Health Physics Society, 78*(6), 744-752.

McGregor, D. K., & Stern, J. A. (1996). Time on task and blink effects on saccade duration. *Ergonomics, 39*(4), 649-660.

Miyake, I., Tange, I., & Hiraga, Y. (1994). MRI findings of the upper eyelid and their relationship with single- and double-eyelid formation. *Aesthetic Plastic Surgery, 18*(2), 183-187.

Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr Eye Res, 21*(1), 535-542.

Morimoto, C. H., Amir, A., & Flickner, M. (2002). *Detecting eye position and gaze from a single camera and 2 light sources.* Paper presented at the Pattern Recognition, 2002. Proceedings. 16th International Conference on.

Morimoto, C. H., Koon, D., Amir, A., & Flickner, M. (2000). Pupil Detection and Tracking Using Multiple Light Sources. *Image and vision computing, 18*(4), 331-335.

Morimoto, C. H., & Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding, 98*(1), 4-24.

Morris, T., Blenkhorn, P., & Zaidi, F. (2002). Blink detection for real-time eye tracking. *Journal of Network and Computer Applications, 25*(2), 129-143.

Morris, T., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology Psychophysiology of Workload, 42*(3), 343-360.

Nguyen, K., Wagner, C., Koons, D., & Flickner, M. (2002). *Differences in the infrared bright pupil response of human eyes.* Paper presented at the Proceedings of the 2002 symposium on Eye tracking research & applications, New Orleans, Louisiana.

Noureddin, B., Lawrence, P. D., & Man, C. F. (2005). A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding, 98*(1), 52-82.

NTSB, U. S. (1995). *Factors that affect fatigue in heavy truck accidents: Case summaries* (No. NTSB Report Number: SS--95--02): U.S. National Transport Safety Board.

Ogilvie, R. D. (2001). The process of falling asleep. *Sleep Medicine, 5*(3), 247-270.

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology, 117*(9), 1885-1901.

OpenCV. (2001). Open Source Computer Vision Library Reference Manual.

Papageorgiou, C. P., Oren, M., & Poggio, T. (1998). *A general framework for object detection.* Paper presented at the Sixth International conference on computer vision, 1998.

Parasuraman, R., & Davies, D. R. (1984). *Varieties of Attention.* Orlando, FL: Academic Press.

Peiris, M. T. R., Jones, R. D., Carroll, G. J., & Bones, P. J. (2004). *Investigation of lapses of consciousness using a tracking task: preliminary results.* Paper presented at the 26th Annual International Conference of the Engineering in Medicine and Biology Society.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., & Bones, P. J. (2006b). *Detecting Behavioral Microsleeps from EEG Power Spectra.* Paper presented at the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., & Bones, P. J. (2006a). Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of Sleep Research, 15*, 291-300.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., Parkin, P. J., Signal, T. L., et al. (2005a). *Identification of Vigilance Lapses using EEG/EOG by Expert Human Raters.* Paper presented at the 27th Annual International Conference of the Engineering in Medicine and Biology Society.

Perez, A., Cordoba, M. L., Gracia, A., & others. (2003). *A Precise Eye-Gaze Detection and Tracking System.* Paper presented at the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen, Czech Republic.

Philip, P., Vervialle, F., Le Breton, P., Taillard, J., & Horne, J. A. (2001). Fatigue, alcohol, and serious road crashes in France: factorial study of national data. *British Medical Journal, 322*, pp. 829-830.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., et al. (2005). *Overview of the face recognition grand challenge.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Raghavachari, R. (2000). *Near-Infrared application in biotechnology* (Vol. 25). Maryland: CRC Press.

Rau, P. S. (1996). *NHTSA's drowsy driver research program fact sheet.* Washington, DC: National Highway Traffic Safety Administration.

Santamaria, J., & Chiappa, K. H. (1987). The EEG of Drowsiness in Normal Adults. *Journal of Clinical Neurophysiology, 4*(327-382).

Seki, M., Shimotani, M., & Nishida, M. (1998). A study of blink detection using bright pupils. *JSAE Review, 19*(1), 58-61.

Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews, 10*(1), 63-76.

Singh, S., & Papanikolopoulos, N. P. (1999). *Monitoring driver fatigue using facial analysis techniques,* 1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, 1999. Proceedings.

Sirohey, S., & Rosenfeld, A. (2001). Eye detection in a face image using linear and nonlinear filters. *Pattern Recognition Society, 34*, 1367-1391.

Sirohey, S., Rosenfeld, A., & Duric, Z. (2002). A method of detecting and tracking irises and eyelids in video. *Pattern Recognition, 35*(6), 1389-1401.

Sliney, D. (2000). ICNIRP Statement on Light Emitting Diodes (LEDs) and Laser Diodes - Implication for Hazard Assessment. *Health Physics, 78*, 744-752.

SMI. (2005, November 2005). *iView X Hi-Speed specification website*. Retrieved April, 2007, from http://www.smi.de/iv/index.html

Smith, P., Shah, M., & Lobo, N. d. V. (2003). Determining driver visual attention with one camera. *IEEE Transactions on Intelligent Transportation Systems, 4*(4), 205-218.

Stutts, J. C., W., W. J., & V., V. B. (1999). Why do people have drowsy driving crashes? Input from drivers who just did. *AAA foundation for Traffic Safety*.

Tian, Y.-l., Kanade, T., & Cohn, J. F. (2000). *Dual-state parametric eye tracking.* Paper presented at the Fourth IEEE International Conference on Automatic Face and Gesture Recognition.

Tian, Z., & Qin, H. (2005). *Real-time driver's eye state detection.* Paper presented at the IEEE International Conference on Vehicular Electronics and Safety.

Tortora, G. J., & Grabowski, S. R. (2003). *Principles of Anatomy and Physiology* (10th Edition ed.). USA: John Wiley & Sons, Inc.

Trosvall, L., & Akerstedt, T. (1987). Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephologram Clinical Neurophysiology, 66*, 502-511.

Ueno, H., Kaneda, M., & Tsukino, M. (1994). *Development of drowsiness detection system.* Paper presented at the Vehicle Navigation and Information Systems Conference.

Van Orden, K. F., Jung, T.-P., & Makeig, S. (1998, 5-9 Oct.). *Eye activity correlates of fatigue during a visual tracking task.* Paper presented at the Human Factors and Ergonomics Society 42nd Annual Meeting.

Van Orden, K. F., Jung, T.-P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology, 52*(3), 221-240.

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features.* Paper presented at the Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Wang, J. G., Sung, E., & Venkateswarlu, R. (2003). *Eye gaze estimation from a single image of one eye.* Paper presented at the 9th IEEE International Conference on Computer Vision, 2003. Proceedings.

Wang, J. S., Knipling, R. R., & Blincco, L. J. (1996). *Motor vehicle crash involvements: A multi-dimensional problem size assessments.* Washington, DC: National Highway Traffic Safety Administration.

Wang, P., Green, M. B., Ji, Q., & Wayman, J. (2005). *Automatic Eye Detection and Its Validation.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Wang, P., & Ji, Q. (2005). *Learning discriminant features for multi-view face and eye detection.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Welch, G., & Gray, B. (2006). *An Introduction to the Kalman Filter*. Chapel Hill: University of North Carolina.

Wierwille, W. W., & Ellsworth, L. A. (1994a). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention, 26*(5), 571-581.

Wierwille, W. W., & Ellsworth, L. A. (1994b). *Research on Vehicle-based Driver Status/performance Monitoring: Development, Validation, and refinement of algorithms for detection of driver drowsiness.*: DOT HS 808 247.

Wu, J., & Zhou, Z.-H. (2003). Efficient face candidates selector for face detection. *Pattern Recognition, 36*(5), 1175-1186.

Yoo, D. H., & Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding, 98*(1), 25-51.

Zheng, Z., Yang, J., & Yang, L. (2005). A robust method for eye features extraction on color image. *Pattern Recognition Letters, 26*, 2252-2261.

Zhou, Z.-H., & Geng, X. (2004). Projection functions for eye detection. *Pattern Recognition, 37*(5), 1049-1056.

Zhu, Z., & Ji, Q. (2004a). Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications, 15*(3), 139-148.

Zhu, Z., & Ji, Q. (2004b). *Real time and non-intrusive driver fatigue monitoring.* Paper presented at the Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems.

Zhu, Z., & Ji, Q. (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding, 98*(1), 124-154.

Zobel, M., Gebhard, A., Paulus, D., Denzler, J., & Niemann, H. (2000). *Robust facial feature localization by coupled features.* Paper presented at the Proceedings of 4th IEEE International Conference onAutomatic Face and Gesture Recognition.